Coalescing random walks on n-block Markov chains

by

Kathleen Lan

Department of Mathematics
Duke University

Date: ________________________
Approved:

Kevin McGoff, Supervisor

Mark Ryser

Honors Thesis submitted in partial fulfillment of the requirements for the degree of
Bachelor of Science in the Department of Mathematics
in the Trinity College of Duke University
2013

Abstract

Fix a discrete-time Markov chain (V, P) with finite state space V and transition matrix P. Let (V_n, P_n) be the Markov chain on n-blocks induced by (V, P), which we call the n-block process associated with the base chain (V, P). We study coalescing random walks on mixing n-block Markov chains in discrete time. In particular, we are interested in understanding the asymptotic behavior of $\mathbb{E}C_n$, the expected coalescence time for (V_n, P_n), as $n \to \infty$. Define the quantity $L = -\log \lambda$, where λ is the Perron eigenvalue of the matrix Q that has entries $Q_{i,j} = P_{i,j}^2$. We prove the existence of four limits and show that all of them are equal to L: $\lim\limits_{n\to\infty} \frac{1}{n} \log \mathbb{E}C_n$, $\lim\limits_{n\to\infty} \frac{1}{n} \log m_n^*$, $\lim\limits_{n\to\infty} \frac{1}{n} \log \bar{m}_n$, and $\lim\limits_{n\to\infty} -\frac{1}{n} \log \Delta_n$, where m_n^* and $\bar{m}_n$ are the maximum and average meeting times for (V_n, P_n) respectively. We establish the inequalities $0 < L \leq h$, where h is the entropy of P, and show that $L = h$ iff P is a measure of maximal entropy. The formulas and bounds for L provide a complete characterization of $\mathbb{E}C_n$ on the exponential scale.

Contents

Abstract **ii**

Acknowledgements **iv**

1 Introduction **1**

2 Definitions **5**

3 Proofs **8**

3.0.1 General bounds for $\mathbb{E}C$ 8

3.0.2 Bounds on $\mathbb{E}C$ and $\mathbb{E}M_{ij}$ for mixing n-block Markov chains . . 14

3.0.3 Proposition 1: $L = -\lim_{n\to\infty} \frac{1}{n} \log \Delta_n$ 17

3.0.4 Proposition 2: $L = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}C_n$ 26

3.0.5 Proposition 3: Formulas for L in terms of meeting times . . . 27

3.0.6 Theorem 2: Bounds for $L = -\log \lambda$ 28

4 Bernoulli Markov chains **34**

Bibliography **44**

Acknowledgements

I would like to thank my mentor, Professor Kevin McGoff, for his invaluable guidance, collaboration, and support. I would also like to acknowledge Professor Kraines, the PRUV program, and the Duke Mathematics Department for providing me the opportunity to work on this research.

1
Introduction

Every Markov chain is associated with a graph that has vertices representing the states and directed edges between states with a non-zero transition probability. A coalescing random walk on a Markov chain is defined as follows. Place a random walker on each vertex of the graph associated with the chain. As time evolves, the walkers move independently according to the transition probabilities until they meet. When two or more walkers meet, or occupy the same state simultaneously, they coalesce (become one walker or cluster) and move together thereafter. The first time that only one walker remains in the system is the coalescence time.

The coalescence time is of interest in its own right and also due to its relationship with an important parameter in another interacting particle system, the voter model. The voter model can be interpreted as describing the evolution of opinions in a social network and has received considerable attention from the "network science" community (see [5] and references therein). Because the coalescing random walk is dual to the voter model, the consensus time of the voter model is equivalent to the coalescence time of the coalescing random walk. Thus, one of the main objectives in the study of coalescing random walks is to determine the expected coalescence time $\mathbb{E}C$.

The expected coalescence time is well-studied for certain types of graphs, such as the torus in $\mathbb{Z}^d$ [4] and r-regular graphs [3]. However, since $\mathbb{E}C$ is often difficult to calculate directly, methods for estimating the order of $\mathbb{E}C$ are often studied instead. Typically, other parameters of the Markov chain are used to bound $\mathbb{E}C$.

In [3], the authors find sharp asymptotic bounds on the order of magnitude for $\mathbb{E}C$ and several other parameters for random walks on r-regular graphs for $r \geq 3$. For a coalescing random walk on an r-regular graph with n vertices, they estimate $\mathbb{E}C$ as $2\theta_r n$, where $\theta_r = \frac{r-1}{r-2}$. Note that $1 < \theta_r \leq 2$. In other words, $\mathbb{E}C$ is approximated as a constant multiple of $|V|$, the cardinality of the state space of the Markov chain.

Hitting times and meeting times (See Definitions 2.0.7 and 2.0.8) have also been used to estimate $\mathbb{E}C$. For transitive, reversible, irreducible, continuous-time Markov chains, Oliveira (2012) proves that $\mathbb{E}C$ is approximately $2m(Q)$ where $m(Q)$ is the expected meeting time of two independent continuous-time random walks [8]. A similar result is also proved for general mixing continuous-time chains which estimates $\mathbb{E}C$ as a constant multiple of the expected meeting time [8]. For reversible, irreducible, continuous-time chains over a finite state space, Oliveira (2010) proves that there exists a universal constant $K > 0$ such that, for any number $n \geq 1$ of independent coalescing random walks, $\mathbb{E}C_n$ has an upper bound of K times the largest expected hitting time on the chain [9].

Although $\mathbb{E}C$ has been studied for certain graphs and for reversible, continuous-time Markov chains, there remain unanswered questions about estimating $\mathbb{E}C$ for non-reversible, discrete-time Markov chains. Our work is focused on characterizing $\mathbb{E}C$ for the discrete time n-block Markov chain (see Definition 2.0.9). The n-block process on a Markov chain (V, P) (see Definition 2.0.2) is the order n version of (V, P). Thus, as $n \to \infty$ the memory of the chain increases. We study the asymptotic behavior of parameters of the n-block process as $n \to \infty$. This can be thought of as the "long-memory limit."

Our main results are stated precisely in the following theorem. For precise definitions of (V_n, P_n) and the stationary distribution π and a formula for entropy h, see Definitions 2.0.9, 2.0.5, and 2.0.13. For convenience, we define some notation which will be used throughout this document. Let (V_n, P_n, π_n) be the n-block Markov chain induced on a mixing Markov chain (V, P, π). Let $\mathbb{E}M_{ij}$ denote the expected meeting time of two random walkers started at states i and j. Define

$$m_n^* = \max_{i,j\in V_n} \mathbb{E}M_{ij},$$
$$\bar{m}_n = \mathbb{E}_{\pi_n\times\pi_n}(\mathbb{E}M_{i,j}),$$
$$\Delta_n = \sum_{u\in V_n} ((\pi_n)_u)^2.$$

Note that m_n^* is the maximum expected meeting time and $\bar{m}_n$ is the average expected meeting time over all pairs of states in V_n. We can interpret Δ_n as follows: Pick two states from V_n randomly and with replacement. Then Δ_n is the probability of picking the same state twice.

Theorem 1.0.1. *Let (V, P) be a non-trivial mixing discrete-time Markov chain with state space V and transition matrix P with entropy $h(V, P)$. Let (V_n, P_n) be the n-block process induced by (V, P). Let π_n be the stationary distribution for (V_n, P_n). Let $\mathbb{E}C_n$ denote the expected coalescence time for (V_n, P_n). Define the matrix Q such that $Q_{ij} = P_{ij}^2$. Let λ be the Perron eigenvalue of Q, and define $L = -\log\lambda$. Then* *the following limits exist and are equal to L:*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}C_n \tag{1.1}$$

$$\lim_{n\to\infty} \frac{1}{n} \log(m_n^*) \tag{1.2}$$

$$\lim_{n\to\infty} \frac{1}{n} \log(\bar{m}_n) \tag{1.3}$$

$$\lim_{n\to\infty} -\frac{1}{n} \log \Delta_n. \tag{1.4}$$

Furthermore,

$$0 < L \leq h(V, P), \tag{1.5}$$

and $L = h$ iff P is a measure of maximal entropy.

To obtain these results, we first develop general bounds for $\mathbb{E}C$ in terms of the maximum expected meeting time (Propositions 3.0.16 and 3.0.20) :

$$\max \mathbb{E}M_{ij} \leq \mathbb{E}C \leq (e+2)(\log|V|+1) \max \mathbb{E}M_{ij}. \tag{1.6}$$

The upper bound in (1.6) was inspired by an analogous result for continuous time chains from [1]. Aldous-Fill [1] provides an upper bound on $\mathbb{E}C$ in terms of the maximum expected hitting time for continuous time Markov chains. We adapt their proof to prove a similar upper bound in terms of the maximum expected meeting time for discrete time Markov chains. For this proof, we also bound the tail of the probability distribution of the meeting time using an exponential term (see Lemma 3.0.18).

In Lemma 3.0.26, we establish a version of subadditivity for Δ_n, which allows us to prove the existence of $\lim_{n\to\infty} -\frac{1}{n} \log \Delta_n$ in Lemma 3.0.27. Applying Lemma 3.0.28 to Lemma 3.0.32 allows us to prove that $L = \lim_{n\to\infty} -\frac{1}{n} \log \Delta_n$ in Lemma 3.0.33. In Corollary 3.0.25 we determine bounds for $\mathbb{E}C_n$ in terms of Δ_n using Corollary 3.0.23 and Proposition 3.0.20. These bounds are used to prove that $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}C_n$ exists and equals L (Lemma 3.0.34). Then, (1.6) is used to prove that $\lim_{n\to\infty} \frac{1}{n} \log(m_n^*)$ exists and equals $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}C_n$, and therefore equals L.

In Lemma 3.0.37, we use the Shannon-McMillan-Breiman theorem and Jensen's inequality to obtain $L \leq h$. We bound the stationary distribution by exponential terms in Lemma 3.0.38, which then allows us to prove $L > 0$ in Lemma 3.0.39. Finally, in Lemma 3.0.40 we use theorems and facts stated in [7] and [2] to prove that $L = h$ iff P is a measure of maximal entropy.

There are a couple of interesting points in the results described in Theorem 1.0.1. First, note that the bounds on L imply that $\mathbb{E}C_n$ (as well as m_n^* and $\bar{m}_n$) is exponential in n. Secondly, the equality of the limits means that $\mathbb{E}C_n$, m_n^* and $\bar{m}_n$ are equal on the exponential scale asymptotically. This result is somewhat surprising, given that full coalescence involves the meeting of many walkers, whereas the meeting time only considers the meeting of two walkers. In the context of the “big bang” idea, however, this result seems reasonable [5]. At the beginning of the coalescing random walk, coalescence events occur quickly as every walker is close to at least one other walker. (Recall that by assumption there are no isolated points in the graph G, so each vertex is associated with at least one edge. Thus, initially, each walker is one edge away from at least one other walker.) It is reasonable that the time required for the last meeting of walkers in the system contributes significantly to the coalescence time. This is easily understood for the meeting of two walkers on a large graph (i.e. for a Markov chain with a large state space) where the distance between the walkers may be significant.

2
Definitions

Definition 2.0.2. *A* **finite-state, discrete time Markov chain** *is defined by*

1. *a finite set V, called the* **state space***;*
2. **a transition matrix** P*, consisting of elements $P_{i,j}$ for $i, j \in V$, where $P_{i,j}$ is the probability of transitioning from state i to state j;*
3. **the Markov property:**
   Let X_t denote the state of the chain at time $t > 0$ and $s_k \in V$ for all $k \geq 1$. Then the Markov property states that

   $$\mathbb{P}(X_t = s_t \mid X_1 = s_1, X_2 = s_2, ..., X_{t-1} = s_{t-1}) = P_{s_{t-1},s_t}.$$

   This is also called the memoryless property.

We denote a Markov chain with state space V and transition matrix P by (V, P).

Definition 2.0.3. *A Markov chain is* **non-trivial** *if $|V| > 1$ (i.e. the Markov chain has more than one state).*

Definition 2.0.4. *A Markov chain (V, P) is* **Bernoulli** *if $\forall\ i, j \in V$,*

$$P_{i,j} = \pi_j$$

where π_j is the stationary probability of state j (see Definition 2.0.5).

Definition 2.0.5. *A* **stationary distribution** π *for a Markov chain (V, P) is a probability vector such that*

$$\pi = \pi P.$$

This vector represents a probability distribution on V. If P is irreducible, then $\pi_j > 0$ for all $j \in V$.
Remark: If (V, P) is mixing, then there exists a unique stationary distribution π.

Definition 2.0.6. *A* **coalescing random walk** *on the Markov chain* (V, P) *is a coupling of* k *Markov processes* (V, P) *such that*

- *for each* i, $\{X_t^i\}$ *is a random walker on the Markov chain* (V, P) *with initial state* $i \in V$. *Let* S *denote the set of initial states of the* k *walkers.*
- *for any* $i, j \in S$, $\{X_t^i\}$ *and* $\{X_t^j\}$ *are independent for* $t \leq \tau_{ij}$, *where* $\tau_{ij} = \inf\{t > 0 | X_t^i = X_t^j\}$.
- *if* $X_t^i = X_t^j$ *for some* t, *then* $X_s^i = X_s^j$ *for all* $s \geq t$.

The **coalescence time** *is given by* $\sup_{i,j \in S}\{\tau_{ij}\}$.

It is often more practical to estimate or bound $\mathbb{E}C$ in terms of other parameters of the Markov chain rather than calculating $\mathbb{E}C$ directly. The hitting time and meeting time parameters are especially useful in this regard.

Definition 2.0.7. *Suppose* X_t *is the observed state at time* t *of a Markov chain with state space* V *and transition matrix* P. *We define the* **hitting time** *of* $j \in V$ *as*

$$T_j = \inf\{t > 0 : X_t = j\}$$

Let $i \in V$ *be the initial state of* X_t. *We write the expected hitting time of* j *from state* i *as* $\mathbb{E}_i T_j$.

Definition 2.0.8. *Let* V *be a finite set and* P *be a probability matrix that define a Markov chain. Let* X_t *and* Y_t *be independent copies of the Markov chain with initial states* i *and* j *respectively. We define the* **meeting time** *as*

$$M_{i,j} = \inf\{t > 0 : X_t = Y_t\}$$

Our work focuses on the families of n-block processes on Markov chains. The concept of n-block processes is introduced in Example 16 of Chapter 14 of [1]. We only consider discrete-time processes, as n-block processes are naturally discrete. The n-block process is defined as follows:

Definition 2.0.9. *Let* (V, P) *be a Markov chain with state space* V *and transition matrix* P. *We call* (V, P) *the base chain or underlying chain. For* $n \geq 1$, *the* **n-block chain** (V_n, P_n) *associated to* (V, P) *has state space* $V_n = \{x = x_1 x_2 ... x_n \mid x_k \in V \text{ for all } k\} \subseteq V^n$. *For* $x \in V_n$ *with* $x = x_1 x_2 ... x_n$, *let* $x[k, l] = x_k ... x_l$. *The transition matrix* P_n *has entries*

$$(P_n)_{i,j} = \begin{cases} P_{i[n],j[n]} & \text{if } i[2, n] = j[1, n-1] \\ 0 & \text{otherwise} \end{cases}$$

Definition 2.0.10. *Let* (V, P) *be a Markov chain with state space* V *and transition matrix* P. *The* **product chain** *on* (V, P) *is the Markov chain with state space* $\{(u, v) \mid u, v \in V\}$ *and transition probabilities* $P_{(i,j),(i^*,j^*)} = P_{i,i^*} P_{j,j^*}$. *We say* (V, P) *is the base chain for the product chain.*

Definition 2.0.11. *The* **diagonal** *of the product chain is defined as*

$$D = \{(v, v) | v \in V\}$$

Note that a meeting time on a Markov chain is equivalent to a hitting time of the diagonal of the corresponding product chain. This is useful, because the hitting time is often easier and simpler to work with than the meeting time.

Definition 2.0.12. *A Markov chain is* **mixing** *if* $\exists k > 0$ *such that* $P_{ij}^k > 0$ *for all* $i, j \in V$.

Formula 2.0.13. *A formula for the* **entropy** *of a mixing Markov chain* (V, P, π) *is given by*

$$h = -\sum_{i,j \in V} \pi_i P_{ij} \log(P_{ij})$$

Entropy is a fundamental quantity of the Markov chain that can be interpreted as a measure of randomness in the Markov chain. In information theory, entropy is equivalent to the information content. Shannon introduced entropy as the expected value of the information contained in a message [10]

Now we will state some fundamental results regarding Markov chains. It can be shown that the mixing property is equivalent to the property of being both irreducible and aperiodic. The Markov chain convergence theorem states that irreducible and aperiodic Markov chains converge to their stationary distribution. This theorem is stated in [6] as follows

Theorem 2.0.14 ([6])**.** *Suppose that* P *is irreducible and aperiodic, with stationary distribution* π. *Then there exists constants* $\alpha \in (0, 1)$ *and* $C > 0$ *such that*

$$\max_{x \in V} \|P^t(x, \cdot) - \pi\|_{TV} \leq C\alpha^t$$

We will also state the following fact.

Fact 2.0.15. *Let* (V, P) *be a mixing Markov chain with finite state space* V *and transition matrix* P. *Let* $\mathbb{E}C$ *be the expected coalescence time for* (V, P). *Then*

$$\mathbb{E}C < \infty.$$

3

Proofs

3.0.1 General bounds for $\mathbb{E}C$

In this section, we bound the expected coalescence time of a general mixing discrete-time Markov chain in terms of the maximum meeting time of the chain.

Proposition 3.0.16. *Let (V, P) be a Markov chain with finite state space V and transition matrix P. Then*

$$\max_{i,j \in V} \mathbb{E}M_{ij} \leq \mathbb{E}C.$$

Proof. Let $i^*, j^* \in V$ be s.t. $\mathbb{E}M_{i^*,j^*} = \max_{i,j} \mathbb{E}M_{ij}$. Label the vertices in a linear order so that $i \prec i^* \prec j^*$ for all $i \in V$ such that $i \neq i^*$ and $i \neq j^*$. Define W_k as the walker started at state k. Construct the coalescing random walk so that, for any $i, j \in V$ such that $i \prec j$, when walkers W_i and W_j meet, they cluster and thereafter follow W_j. By this construction, we have

$$M_{i^*j^*} \leq C. \tag{3.1}$$

That is, a walker that meets with W_{i^*} or W_{j^*} thereafter follows W_{i^*} or W_{j^*} respectively. There are two cases: either the cluster containing i^* has already coalesced with the cluster containing j^* before the coalescence time C or the coalescence of the two clusters results in full coalescence. (We see that $M_{i^*j^*} > C$ implies that the cluster containing i^* has not yet coalesced with the cluster containing j^* at the full coalescence time. This is a contradiction of the definition of C.) These two cases imply (3.1) as desired. Taking the expectation of both sides of (3.1) completes the proof. □

To obtain the upper bound, we first need to prove a few lemmas. We will adapt a proof from Chapter 2 (pg. 19) of [1] to prove the following lemma.

Lemma 3.0.17. *Let (V, P) be a discrete-time Markov chain with finite state space V and transition matrix P. Let T_B be the first hitting time on $B \subset V$, and let $t_B = \max_i \mathbb{E}_i T_B$. For any $i, j \in V$,*

$$\mathbb{P}_\mu(T_B > t) \leq e^{-\frac{t}{(e+1)t_B}}. \tag{3.2}$$

Proof. For any initial distribution μ, any integer time $s > 0$, and any integer $m \geq 1$,

$$\begin{aligned}\mathbb{P}_\mu(T_B > ms | T_B > (m-1)s) &= \mathbb{P}_\theta(T_B > s) \text{ for some distribution } \theta \text{ that depends on } \mu \\ &\leq \max_i \mathbb{P}_i(T_B > s).\end{aligned} \tag{3.3}$$

Regarding notation: $\mathbb{P}_i(T_B > s)$ is the probability that $T_B > s$ given an initial state i.

Applying Markov's inequality to the right-hand side of (3.3) yields

$$\mathbb{P}_\mu(T_B > ms | T_B > (m-1)s) \leq \frac{\max_i \mathbb{E}_i T_B}{s} = \frac{t_B}{s}. \tag{3.4}$$

Now we will use mathematical induction on m to show that

$$\mathbb{P}_\mu(T_B > ms) \leq \left(\frac{t_B}{s}\right)^m \text{ for } m \geq 1. \tag{3.5}$$

We start by considering (3.4) for $m = 1$:

$$\mathbb{P}_\mu(T_B > s | T_B > 0) = \mathbb{P}_\mu(T_B > s) \leq \frac{t_B}{s}.$$

We see that (3.5) is true for $m = 1$, so for some m the following inequality holds:

$$\mathbb{P}_\mu(T_B > ms) \leq \left(\frac{t_B}{s}\right)^m. \tag{3.6}$$

By (3.4),

$$\mathbb{P}_\mu(T_B > (m+1)s | T_B > ms) \leq \frac{t_B}{s}. \tag{3.7}$$

Multiplying both sides of (3.7) by $\mathbb{P}_\mu(T_B > ms)$ gives

$$\begin{aligned}\mathbb{P}_\mu((T_B > (m+1)s) \cap (T_B > ms)) = \mathbb{P}_\mu(T_B > (m+1)s | T_B > ms)\mathbb{P}_\mu(T_B > ms) &\leq \frac{t_B}{s}\mathbb{P}_\mu(T_B > ms) \\ &\leq \left(\frac{t_B}{s}\right)^{m+1} \text{ by (3.6).}\end{aligned}$$

In general, for a random variable X, and constants c_1 and $c_2 > 0$,

$$\mathbb{P}((X > c_1 + c_2) \cap (X > c_1)) = \mathbb{P}(X > c_1 + c_2).$$

So our inequality simplifies to

$$\mathbb{P}_\mu(T_B > (m+1)s) \le \left(\frac{t_B}{s}\right)^{m+1}.$$

Thus, by mathematical induction on the positive integers, we obtain (3.5) as desired.

Let $t = js$ and substitute into (3.5):

$$\mathbb{P}_\mu(T_B > t) \le \left(\frac{t_B}{s}\right)^{\frac{t}{s}}. \tag{3.8}$$

Note that (3.8) holds for any $s \in \mathbb{N}$, but we would like to optimize this bound by finding an s that minimizes $f(s,t) = \left(\frac{t_B}{s}\right)^{\frac{t}{s}}$. Although s must be an integer, we will first minimize f over all real numbers $s > 0$ as doing so will make it clear which integer value of s minimize f.

First, differentiate f with respect to s:

$$\begin{aligned}\frac{d}{ds}\left[\left(\frac{t_B}{s}\right)^{\frac{t}{s}}\right] &= \frac{d}{ds}\left[e^{\frac{t}{s}\ln\left(\frac{t_B}{s}\right)}\right] \\ &= e^{\frac{t}{s}\ln\left(\frac{t_B}{s}\right)}\left(-\frac{t}{s^2}\ln\left(\frac{t_B}{s}\right) + \frac{t}{s}\left(\frac{s}{t_B}\right)\left(-\frac{t_B}{s^2}\right)\right) \\ &= e^{\frac{t}{s}\ln\left(\frac{t_B}{s}\right)}\left(-\frac{t}{s^2}\ln\left(\frac{t_B}{s}\right) - \frac{t}{s^2}\right) \\ &= e^{\frac{t}{s}\ln\left(\frac{t_B}{s}\right)}\left(-\frac{t}{s^2}\right)\left(\ln\left(\frac{t_B}{s}\right)+1\right).\end{aligned} \tag{3.9}$$

The first two factors on the right-hand side of (3.9) are always non-zero, so we only need to consider the third factor. Let $\ln\left(\frac{t_B}{s}\right) + 1 = 0$. Then,

$$\begin{aligned}\ln\left(\frac{t_B}{s}\right) &= -1 \\ \frac{t_B}{s} &= e^{-1} \\ s &= et_B.\end{aligned}$$

We don't have any boundary points to check, so $s = et_B$ is our only critical point.

Now let us conduct a first derivative test. Conveniently,

$$2t_B \le et_B \le 3t_B.$$

We find that

$$f_s(2t_B, t) < 0 \text{ and } f_s(3t_B, t) > 0.$$

Thus, the first derivative test tells us that f attains a minimum at $s = et_B$. Since f is strictly decreasing for $s < et_B$ and strictly increasing for $s > et_B$, the integer value of s that minimizes f must be one of the closest integers on either side of $s = et_B$. That is, this integer is either the largest integer smaller than $s = et_B$, i.e. $\lfloor et_B \rfloor$, or the smallest integer larger than $s = et_B$, i.e. $\lceil et_B \rceil$.

There doesn't seem to be an obvious or simple way to determine which of $\lfloor et_B \rfloor$ and $\lceil et_B \rceil$ is the more optimal choice with regards to minimizing $f(s,t)$, but we can still use either one to bound $\mathbb{P}(T_B > t)$. It turns out that choosing $s = \lceil et_B \rceil$ makes things easier later on, so we select $s = \lceil et_B \rceil$.

Our inequality then becomes

$$\mathbb{P}_\mu(T_B > t) \leq \left(\frac{t_B}{\lceil et_B \rceil} \right)^{\frac{t}{\lceil et_B \rceil}}. \tag{3.10}$$

If we let k be a constant defined as $k = \frac{\lceil et_B \rceil}{t_B}$, then we can write $\lceil et_B \rceil = kt_B$. Note that $e \leq k \leq e+1$ (since $et_B \leq \lceil et_B \rceil = kt_B \leq (e+1)t_B$ and $t_B > 0$), and so $\frac{1}{e+1} \leq \frac{1}{k} \leq \frac{1}{e}$. Now we can rewrite (3.10) and finish the proof.

$$\begin{aligned} \mathbb{P}_\mu(T_B > t) &\leq \left(\frac{t_B}{kt_B} \right)^{\frac{t}{kt_B}} \\ &= \left(\frac{1}{k} \right)^{\frac{t}{kt_B}} \\ &\leq \left(\frac{1}{e} \right)^{\frac{t}{kt_B}} = e^{-\frac{t}{kt_B}} \\ &\leq e^{-\frac{t}{(e+1)t_B}}. \end{aligned} \tag{3.11}$$

□

Corollary 3.0.18. *Let (V, P) be a discrete-time Markov chain with finite state space V and transition matrix P. Let $m^* = \max_{i,j \in V} \mathbb{E} M_{i,j}$. For any $i, j \in V$,*

$$\mathbb{P}(M_{i,j} > t) \leq e^{-\frac{t}{(e+1)m^*}}. \tag{3.12}$$

Proof. Consider the product chain on (V, P). This is the Markov chain with state space $V \times V$ and transition matrix $P \times P$. Let $Z_t = (X_t, Y_t)$ be a copy of the product chain Markov process of independent processes X_t and Y_t, each of which is a Markov process started at states i and j respectively on our original Markov chain. We can view the meeting time on the original Markov chain, $M_{i,j}$, as the hitting time of $D = \{(x, x) : x \in V\}$ on the product chain. Apply Lemma 3.0.17 to the product chain by letting $B = D$. Then $T_D = M_{i,j}$ and $t_D = \max_{i,j} \mathbb{E}_{(i,j)} T_D = \max_{i,j} \mathbb{E} M_{i,j} = m^*$. Thus,

$$\mathbb{P}(M_{i,j} > t) \leq e^{-\frac{t}{(e+1)m^*}}. \tag{3.13}$$

(In this case, μ is the distribution for which state (i,j) has probability 1 and all other states have probability 0.)

□

Next, we will prove the following lemma using calculus.

Lemma 3.0.19. *Let A and a be constants such that $A \geq 1$ and $a > 0$. Then,*

$$\sum_{t=0}^{\infty} \min(1, Ae^{-at}) \leq a^{-1}(1 + \log(A)) + 1. \tag{3.14}$$

Proof. Let $f(t) = \min(1, Ae^{-at})$. For all A and a, if we define $t^* = \frac{\ln(A)}{a}$, then $Ae^{-at^*} = 1$. We know that Ae^{-at} is monotonically decreasing. Therefore, $Ae^{-at} > 1$ for $t < t^*$ and $Ae^{-at} < 1$ for $t > t^*$. Now we can conclude that $f(t) = 1$ for $t \leq t^*$ and $f(t) = Ae^{-at}$ for $t > t^*$, so $f(t)$ is monotonically decreasing as well. This then implies that the right-hand Riemann sum of $f(t)$ on the interval $(0, \infty)$ underestimates the integral of $f(t)$ on the same interval. If we let $\Delta x = 1$, the right-hand Riemann sum is

$$\sum_{t=1}^{\infty} \min(1, Ae^{-at}),$$

and so,

$$\sum_{t=1}^{\infty} \min(1, Ae^{-at}) \leq \int_{0}^{\infty} \min(1, Ae^{-at}) dt.$$

Using calculus,

$$\int_{0}^{\infty} \min(1, Ae^{-at}) dt = a^{-1}(1 + \log(A)),\ A \geq 1,$$

so

$$\sum_{t=1}^{\infty} \min(1, Ae^{-at}) \leq a^{-1}(1 + \log(A)).$$

Add 1 to each side to obtain

$$\sum_{t=0}^{\infty} \min(1, Ae^{-at}) \leq a^{-1}(1 + \log(A)) + 1.$$

□

We will use Lemmas (3.0.17) and (3.0.19) to prove the following upper bound on $\mathbb{E}C$. This proof is inspired by a proof in Chapter 14 (pg. 14-15) of [1].

Proposition 3.0.20. *Consider a general discrete-time Markov chain with state space V. Let $m^* = \max_{i,j \in V} \mathbb{E}M_{ij}$. We have the following upper bound on $\mathbb{E}C$:*

$$\mathbb{E}C \leq (e + 2)(\log|V| + 1)m^*.$$

Proof. We will consider the coalescing random walk (CRW) of walkers of a general Markov process on a graph with vertex set V. First, label the vertices 1, 2, ..., $|V|$. Construct the CRW so that vertices that meet with vertex 1 thereafter follow the path of vertex 1. This construction yields

$$C \leq \max_j M_{1,j}.$$

Note that this implies

$$\mathbb{P}(C > t) \leq \mathbb{P}(\max_j M_{1,j} > t) \leq \sum_j \mathbb{P}(M_{1,j} > t). \tag{3.15}$$

Now, consider $\mathbb{E}C$ by using the Tail-Sum formula for expectation,

$$\mathbb{E}C = \sum_{t=0}^{\infty} \mathbb{P}(C > t) \tag{3.16}$$

$$\leq \sum_{t=0}^{\infty} \min(1, \sum_j \mathbb{P}(M_{1,j} > t)) \tag{3.17}$$

$$\tag{3.18}$$

We see that (3.17) results from (3.15) and the fact that $\mathbb{P}(C > t) \leq 1$ for all t. By Lemma 3.0.17, we obtain

$$\mathbb{E}C \leq \sum_{t=0}^{\infty} \min(1, |V| e^{-\frac{t}{(e+1)m^*}}) \tag{3.19}$$

where $m^* = \max_{i,j \in V} EM_{ij}$. Apply Lemma 3.0.19 to the right-hand side of (3.19) to finish the proof:

$$\mathbb{E}C \leq (e+1)m^*(1 + \log|V|) + 1 \leq (e+2)(\log|V| + 1)m^*. \tag{3.20}$$

□

Corollary 3.0.21. *For general mixing discrete-time Markov chains (V, P) with finite state space V and transition matrix P,*

$$\max_{i,j \in V} \mathbb{E}M_{ij} \leq \mathbb{E}C \leq (e+2)(\log|V| + 1) \max_{i,j \in V} \mathbb{E}M_{ij}. \tag{3.21}$$

Proof. This corollary follows from Proposition 3.0.16 and Proposition 3.0.20. □

3.0.2 Bounds on $\mathbb{E}C$ and $\mathbb{E}M_{ij}$ for mixing n-block Markov chains

Lemma 3.0.22. *Let (V, P) be a non-trivial discrete-time mixing Markov chain with finite state space V and transition matrix P. Let (V_n, P_n) be the n-block process that arises from (V, P). Define $\Delta_n = \sum_{u \in V_n} (\pi_n)_u^2$. Then, $\exists k_1 > 0$ such that, for $i, j \in V_n$ chosen from π_n,*

$$\mathbb{E}M_{i,j} \geq \frac{k_1}{\Delta_n}.$$

We see that the following corollary follows directly from the lower bound in (3.21) and Lemma 3.0.22.

Corollary 3.0.23. *In the setting of Lemma 3.0.22, it holds that*

$$\mathbb{E}C_n \geq \frac{k_1}{\Delta_n}.$$

Proof of Lemma 3.0.22. Let $i, j \in V_n$ be chosen from π_n. Let X and Y be independent copies of the n-block Markov chain with states at time t described by X_t and Y_t respectively. We have

$$\mathbb{P}(M_{ij} \leq k) = \mathbb{P}(\bigcup_{g=0}^{k-1} \{X_{1+g}^{n+g} = Y_{1+g}^{n+g}\}) \leq k\mathbb{P}(X_1^n = Y_1^n) = k\Delta_n, \tag{3.22}$$

where the inequality is a consequence of the union bound and of the stationarity of the chain.

Define $K = \sup\{t \mid t\Delta_n \leq 1\} = \lfloor \frac{1}{\Delta_n} \rfloor$. The following equalities and inequalities hold:

$$
\begin{aligned}
\mathbb{E}M_{ij} &= \sum_{t=0}^{\infty} \mathbb{P}(M_{ij} > t) \\
&= \sum_{t=0}^{\infty} (1 - \mathbb{P}(M_{ij} \leq t)) \\
&\geq \sum_{t=0}^{K} (1 - \mathbb{P}(M_{ij} \leq t)) \\
&\geq \sum_{t=0}^{K} (1 - t\Delta_n) \text{ by (3.22)} \\
&= (K+1) - \Delta_n K \left(\frac{K+1}{2} \right) \\
&\geq (K+1)(1 - \frac{\Delta_n K}{2}) \\
&\geq \frac{1}{2}(K+1) \\
&\geq \frac{k_1}{\Delta_n}.
\end{aligned}
$$

□

Lemma 3.0.24. *Let (V, P) describe a mixing discrete-time Markov chain with state space V and transition matrix P. Let (V_n, P_n) be the n-block process on (V, P). Define $\Delta_n = \sum_{u \in V_n} (\pi_u)^2$. Then $\exists k_1, k_2 > 0$ such that $\forall\ i, j \in V_n$,*

$$k_1 \frac{1}{\Delta_n} \leq \mathbb{E}M_{ij} \leq k_2 n \frac{1}{\Delta_n}.$$

Proof. The lower bound is Corollary 3.0.22.

Consider the transition matrix for the product chain of (V, P), which we will denote as P_*. Because V is mixing, for any $a, b, c \in V$, we have that $\frac{P_*^T((a,b),(c,c))}{\pi_c^2} \to 1$ as $T \to \infty$. Thus, for any $\epsilon > 0$, $\exists T$ such that for all $x[n], y[n], u[1] \in V$, $\frac{P_*^T((x[n],y[n]),(u[1],u[1]))}{\pi_{u[1]}^2} \geq 1 - \epsilon$.

Fix $0 < \epsilon < 1$ and the corresponding T. Let X_t^i denote the n-block state at time $t > 0$ of a random walker started at state $i \in V_n$ at $t = 0$. Then for arbitrary $i, j, u \in V_n$,

$$\mathbb{P}(X_{n+T}^i = X_{n+T}^j = u) = P_*^T((i[n], j[n]), (u[1], u[1])) \frac{\pi_u^2}{\pi_{u[1]}^2} \geq (1 - \epsilon)\pi_u^2. \tag{3.23}$$

Summing (3.23) over all $u \in V_n$ gives

$$\mathbb{P}(M_{ij} \leq n + T) \geq \mathbb{P}(X_{n+T}^i = X_{n+T}^j) \geq (1 - \epsilon) \sum_{u \in V_n} \pi_u^2 = (1 - \epsilon)\Delta_n. \tag{3.24}$$

Next, we will use mathematical induction to prove $\mathbb{P}(M_{ij} > m(n+T)) \leq (1-(1-\epsilon)\Delta_n)^m$ for m ≥ 1. Consider the base case, m = 1. Trivially, by (3.24), we have

$$\mathbb{P}(M_{ij} > n+T) = 1 - \mathbb{P}(M_{ij} \leq n+T) \leq 1-(1-\epsilon)\Delta_n.$$

Now suppose $\mathbb{P}(M_{ij} > m(n+T)) \leq (1-(1-\epsilon)\Delta_n)^m$ for some m. We can express $\mathbb{P}(M_{ij} > (m+1)(n+T))$ as follows:

$$\mathbb{P}(M_{ij} > (m+1)(n+T)) = \mathbb{P}(M_{ij} > (m+1)(n+T) | M_{ij} > m(n+T)) \cdot \mathbb{P}(M_{ij} > m(n+T)).$$

We define the notation $i^* := X^i_{m(n+T)}$ and $j^* := X^j_{m(n+T)}$. We can write

$$\mathbb{P}(M_{ij} > (m+1)(n+T)) = \mathbb{P}(M_{i^*j^*} > (n+T)) \cdot \mathbb{P}(M_{ij} > m(n+T)).$$

By (3.24) and our hypothesis,

$$\mathbb{P}(M_{ij} > (m+1)(n+T)) \leq (1-(1-\epsilon)\Delta_n) \cdot (1-(1-\epsilon)\Delta_n)^m = (1-(1-\epsilon)\Delta_n)^{m+1}.$$

By induction on m, we prove that $\mathbb{P}(M_{ij} > m(n+T)) \leq (1-(1-\epsilon)\Delta_n)^m$ for all m ≥ 1.

Now define a geometric random variable G with parameter $(1-\epsilon)\Delta_n$. Note that we can write $\mathbb{P}(G > m) = (1-(1-\epsilon)\Delta_n)^m$. Thus,

$$\mathbb{P}(M_{ij} > m(n+T)) = \mathbb{P}\left(\frac{M_{ij}}{n+T} > m\right) \leq \mathbb{P}(G > m).$$

We can sum over all m and use the Tail-Sum formula to obtain

$$\mathbb{E}\left(\frac{M_{ij}}{n+T}\right) \leq \mathbb{E}G = \frac{1}{(1-\epsilon)\Delta_n}.$$

Since n+T is a constant, we have

$$\frac{\mathbb{E}M_{ij}}{n+T} \leq \frac{1}{(1-\epsilon)\Delta_n},$$

which implies

$$\mathbb{E}M_{ij} \leq (n+T)\frac{1}{(1-\epsilon)\Delta_n}.$$

Since T is fixed, we have $n+T \leq Kn$ for some $K \in \mathbb{N}$. Let $k_2 = \frac{K}{1-\epsilon}$. Then we have

$$\mathbb{E}M_{ij} \leq k_2 n \frac{1}{\Delta_n},$$

as desired. $\square$

Corollary 3.0.25. *Under the conditions of Lemma 3.0.24, there exists constants $k_1, k_2 > 0$ such that*

$$k_1 \frac{1}{\Delta_n} \leq \mathbb{E}C_n \leq k_2 n^2 \frac{1}{\Delta_n}.$$

Proof. The lower bound is simply Corollary (3.0.23), so we will show the upper bound.

Apply the upper bound in Proposition 3.0.20 to (V_n, P_n):

$$\mathbb{E}C_n \leq (e+2)(n \log |V| + 1) m_n^* \leq (e+2)(\log |V| + 1) n m_n^*,$$

where $m_n^* = \max_{i,j \in V_n} \mathbb{E}M_{i,j}$. Letting $K = (e+2)(\log |V| + 1)$, we have

$$\mathbb{E}C_n \leq K n m_n^*. \tag{3.25}$$

Apply Lemma 3.0.24 for $i^*, j^* \in V_n$ such that $\mathbb{E}M_{i^*j^*} = \max_{ij} \mathbb{E}M_{ij} = m_n^*$. Then,

$$\begin{aligned} \mathbb{E}C_n &\leq Kn(K_2 n \frac{1}{\Delta_n}) \\ &= k_2 n^2 \frac{1}{\Delta_n}. \end{aligned}$$

where $k_2 = KK_2$. □

3.0.3 Proposition 1: $L = -\lim_{n\to\infty} \frac{1}{n} \log \Delta_n$

Lemma 3.0.26. *Let (V, P, π) describe a discrete time Markov chain with finite state space V, transition matrix P, and stationary distribution π. Let (V_n, P_n, π_n) be the n-block Markov process induced by (V, P). Define $\Delta_n = \sum_{j \in V_n} (\pi_n)_j^2$. Then $\exists k_1, k_2 > 0$ such that*

$$k_1 \Delta_m \Delta_n \leq \Delta_{m+n} \leq k_2 \Delta_m \Delta_n. \tag{3.26}$$

Let $k' = \inf\{t > 0 \mid (P^t)_{ij} > 0 \ \forall i, j \in V\}$, and let $s \geq k'$. Then there exists a constant k_3 such that

$$\Delta_{m+n+s} \geq k_3 \Delta_m \Delta_n. \tag{3.27}$$

Proof. We begin by writing out formulas for $\Delta_m \Delta_n$ and Δ_{m+n}.

$$\Delta_m \Delta_n = \sum_{u \in V_m, w \in V_n} \pi_u^2 \pi_w^2$$

and

$$\Delta_{m+n} = \sum_{u \in V_{m+n}} \pi_u^2 = \sum_{v \in V_m, w \in V_n} \left(\pi_v P_{v[m], w[1]} \frac{\pi_w}{\pi_{w[1]}} \right)^2.$$

Let $m_3 = \min_{w \in V} \pi_w$ and $k_2 = \left(\frac{1}{m_3}\right)^2$. Then we can obtain the upper bound in (3.26):

$$\begin{aligned}\Delta_{m+n} &\leq \sum_{u \in V_m, w \in V_n} {\pi_u}^2 {\pi_w}^2 \left(\frac{1}{\pi_{w[1]}}\right)^2 \\ &\leq k_2 \sum_{u \in V_m, w \in V_n} {\pi_u}^2 {\pi_w}^2 = k_2 \Delta_m \Delta_n.\end{aligned}$$

Fix $s \geq k'$. Now we prove that $\Delta_{m+n+s} \geq k_3 \Delta_m \Delta_n$ for some constant k_3.

We first write a formula for Δ_{m+n+s}:

$$\Delta_{m+n+s} = \sum_{v \in V_m, u \in V_s, w \in V_n} \left(\pi_v P_{v[m],u[1]} \frac{\pi_u}{\pi_{u[1]}} P_{u[s],w[1]} \frac{\pi_w}{\pi_{w[1]}}\right)^2$$

By the mixing property of P and the constraint on s, for all $v, w \in V$, $\exists u \in V_s$ such that $\mathbb{P}(X_1^{m+s+n} = vuw) > 0$. Let

$$k_3 = \min_{\mathbb{P}(X_1^{m+s+n} = vuw) > 0} \left(\frac{P_{vu[1]}}{\pi_{u[1]}} \pi_u \frac{P_{u[s]w}}{\pi_w}\right)^2 > 0.$$

Then,

$$\Delta_{m+n+s} \geq k_3 \sum_{v \in V_m, u \in V_s, w \in V_n} (\pi_v \pi_w)^2 = k_3 \Delta_m \Delta_n$$

which proves (3.27).

Now to show the lower bound, we will use the two results proved above. From the upper bound in (3.26), we have

$$\begin{aligned}\Delta_{m+n+s} &\leq k_2 \Delta_{m+n} \Delta_s \\ &\leq k_2 \Delta_{m+n} \quad \text{since } \Delta_s \leq 1.\end{aligned}$$

Combining the inequality above with (3.27), we have

$$k_3 \Delta_m \Delta_n \leq \Delta_{m+n+s} \leq k_2 \Delta_{m+n},$$

which can be rewritten as

$$\frac{k_3}{k_2} \Delta_m \Delta_n \leq \Delta_{m+n}.$$

Let $k_1 = \frac{k_3}{k_2}$ and we are done. □

Now we will use Lemma 3.0.26 to prove that $\lim_{n \to \infty} -\frac{1}{n} \log(\Delta_n)$ exists. The proof is very similar to the proof for the subadditivity lemma [7].

Lemma 3.0.27. *Let (V, P, π) describe a discrete time Markov chain with finite state space V, transition matrix P, and stationary distribution π. Let (V_n, P_n, π_n) be the n-block Markov process induced by (V, P). Define $\Delta_n = \sum_{j \in V_n} (\pi_n)_j^2$. Then, $\lim_{n\to\infty} -\frac{1}{n}\log(\Delta_n)$ exists.*

Proof. Let $n = qm + r$, $0 \le r < m$ for any $n, m > 0$. Then by Lemma (3.0.26),

$$\Delta_n = \Delta_{qm+r} \ge k_1 \Delta_{qm} \Delta_r. \tag{3.28}$$

Using Lemma 3.0.26 q times, we obtain

$$\Delta_{qm} \ge (k_1)^{q-1} (\Delta_m)^q. \tag{3.29}$$

By (3.28) and (3.29),

$$\Delta_n \ge (k_1)^q (\Delta_m)^q \Delta_r.$$

Taking the logarithm yields

$$\log(\Delta_n) \ge q\log(k_1) + q\log(\Delta_m) + \log(\Delta_r).$$

Multiplying by $\frac{-1}{n}$ gives

$$\left(-\frac{1}{n}\right)\log(\Delta_n) \le \frac{-q\log(k_1)}{n} - \frac{q\log(\Delta_m)}{n} - \frac{\log(\Delta_r)}{n}.$$

Note that $q = \frac{n-r}{m}$. Let $k = \min_{0\le r<m} \log(\Delta_r)$. Then,

$$\begin{aligned}\frac{-\log(\Delta_n)}{n} &\le \frac{-(\frac{n-r}{m})\log(k_1)}{n} - \left(\frac{n-r}{n}\right)\frac{\log(\Delta_m)}{m} - \frac{k}{n} \\ &= \left(-\frac{1}{m} + \frac{r}{nm}\right)\log(k_1) - \left(\frac{n-r}{n}\right)\frac{\log(\Delta_m)}{m} - \frac{k}{n}.\end{aligned}$$

Taking the limit supremum in n yields

$$\limsup_n \frac{-\log(\Delta_n)}{n} \le -\frac{1}{m}\log(k_1) - \frac{\log(\Delta_m)}{m}.$$

Taking the limit infimum in m yields:

$$\limsup_n \frac{-\log(\Delta_n)}{n} \le \liminf_m \frac{-\log(\Delta_m)}{m} = \liminf_n \frac{-\log(\Delta_n)}{n}.$$

We also always have that $\liminf -\frac{1}{n}\log(\Delta_n) \le \limsup -\frac{1}{n}\log(\Delta_n)$. Thus,

$$\liminf_n -\frac{1}{n}\log(\Delta_n) = \limsup_n -\frac{1}{n}\log(\Delta_n),$$

and so $\lim_{n\to\infty} -\frac{1}{n}\log(\Delta_n)$ exists. $\square$

Next, we prove that $L = \lim_{n\to\infty} -\frac{1}{n}\log \Delta_n$ using the following lemmas.

Lemma 3.0.28. *Let $\{a_n\}$ be a sequence of positive real numbers such that $\{a_n\}$ converges to $a > 0$ (i.e. $\lim_{n\to\infty} a_n = a$). Then, $\frac{1}{n}\log \prod_{k=1}^{n} a_k \to \log a$.*

Proof. First, fix $\epsilon > 0$. By the continuity and monotonicity of the logarithm function, $a_n \to a$ implies $\log a_n \to \log a$. Thus, we can choose a constant M_ϵ such that for all $k > M_\epsilon$,

$$|\log(a_k) - \log(a)| < \frac{\epsilon}{2}. \tag{3.30}$$

Now, let us consider the term $\log \prod_{k=1}^{n} a_k$. Note that $\log \prod_{k=1}^{n} a_k = \sum_{k=1}^{n} \log a_k$ by a property of logarithms. For sufficiently large n (i.e. for $n > M_\epsilon$), we can write

$$\log \prod_{k=1}^{n} a_k = \sum_{k=1}^{M_\epsilon} \log a_k + \sum_{k=M_\epsilon+1}^{n} \log a_k$$

Applying the Triangle Inequality produces the following inequality.

$$\left|\frac{1}{n}\sum_{k=1}^{M_\epsilon} \log(a_k) + \frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \log(a_k) - \log(a)\right| \le \frac{1}{n}\sum_{k=1}^{M_\epsilon}\left|\log(a_k)\right| + \left|\frac{1}{n}\left(\sum_{k=M_\epsilon+1}^{n} \log(a_k)\right) - \log(a)\right|. \tag{3.31}$$

We can bound the second term of the right-hand side of (3.31) as follows. First, write

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \log(a_k) - \log(a)\right| &\le \left|\frac{1}{n}\sum_{k=M_\epsilon+1}^{n} [\log(a_k) - \log(a)]\right| \\ &\le \frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \left|\log(a_k) - \log(a)\right| \text{ by the Triangle Inequality.}\end{aligned}$$

Then, by (3.30),

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \log(a_k) - \log(a)\right| &\le \frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \frac{\epsilon}{2} \\ &= \frac{n - M_\epsilon}{n}\left(\frac{\epsilon}{2}\right) \\ &< \frac{\epsilon}{2}.\end{aligned}$$

Now, (3.31) becomes

$$\left|\frac{1}{n}\sum_{k=1}^{M_\epsilon} \log(a_k) + \frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \log(a_k) - \log(a)\right| \le \frac{1}{n}\sum_{k=1}^{M_\epsilon} |\log(a_k)| + \frac{\epsilon}{2}. \tag{3.32}$$

Let $M = \max_{1 \leq k \leq M_\epsilon} |\log(a_k)|$. Then we can bound the first term of the right-hand side of (3.32) in terms of M:

$$\frac{1}{n}\sum_{k=1}^{M_\epsilon} |\log(a_k)| \leq \frac{1}{n}\sum_{k=1}^{M_\epsilon} M = \frac{MM_\epsilon}{n}.$$

For sufficiently large n (i.e. for $n > \frac{2MM_\epsilon}{\epsilon}$), we have $\frac{MM_\epsilon}{n} < \frac{\epsilon}{2}$, in which case (3.32) becomes

$$\left|\frac{1}{n}\sum_{k=1}^{M_\epsilon} \log(a_k) + \frac{1}{n}\sum_{k=M_\epsilon+1}^{n} \log(a_k) - \log(a)\right| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, for any $\epsilon > 0$ and sufficiently large n,

$$\left|\frac{1}{n}\sum_{k=1}^{n} \log(a_k) - \log(a)\right| \leq \epsilon.$$

This implies that $\frac{1}{n}\log\prod_{k=1}^{n} a_k \to \log(a)$ as desired. $\square$

At this point, we will make the following definitions.

Let (V, P) be a discrete-time mixing Markov chain with finite state space V and transition matrix P. Let X_t be the state of the product chain $(V, P) \times (V, P)$ at time t. Let A_t be the event that $X_t = (u, u)$ for some $u \in V$, and let $I_t = \cap_{k=0}^{t} A_k$. Then we define the function $g_k(u) = \mathbb{P}(X_k = (u, u), I_k)$ for $k \geq 0$ and $u \in V$. Observe that

$$\begin{aligned} g_{k+1}(v) &= (X_{k+1} = (v, v), I_{k+1}) \\ &= \sum_{u \in V} \mathbb{P}(X_{k+1} = (v, v), A_{k+1} | X_k = (u, u), I_k)\mathbb{P}(X_k = (u, u), I_k) \\ &= \sum_{u \in V} g_k(u)\mathbb{P}(X_{k+1} = (v, v), A_{k+1} | X_k = (u, u), I_k). \end{aligned}$$

Note that if $X_{k+1} = (v, v)$ occurs, then by definition A_{k+1} also occurs with probability 1. Thus, we have

$$\mathbb{P}(X_{k+1} = (v, v), A_{k+1} | X_k = (u, u), I_k) = \mathbb{P}(X_{k+1} = (v, v) | X_k = (u, u), I_k).$$

By the Markov property, we can further reduce this probability as follows.

$$\mathbb{P}(X_{k+1} = (v, v) | X_k = (u, u), I_k) = \mathbb{P}(X_{k+1} = (v, v) | X_k = (u, u)) = P_{uv}^2.$$

Thus, we have the following recursive formula for the family of functions $\{g_n\}$:

$$g_{k+1}(v) = \sum_{u \in V} g_k(u) P_{uv}^2.$$

Now, we define the family of row vectors $\{\mu_k\}$ such that $\mu_k = g_k$. (Note that $|\mu_k| = |V|$.) We also define $S(\mu_k) = \sum_{v\in V} \mu_k(v)$. Observe that

$$S(\mu_k) = \sum_{v\in V} \mathbb{P}(X_k = (v, v), I_k) = \mathbb{P}(I_k).$$

From the definition, we can determine

$$\mu_0(u) = g_0(u) = \mathbb{P}(X_0 = (u, u), I_0) = \mathbb{P}(X_0 = (u, u)) = \pi_u^2.$$

For $k \geq 0$ and $v \in V$,

$$\mu_{k+1}(v) = \sum_{u\in V} \mu_k(u) P_{u,v}^2. \tag{3.33}$$

Now, we define the matrix Q with entries $Q_{ij} = P_{ij}^2$. Since Q is non-negative and $Q_{ij} > 0$ iff $P_{ij} > 0$, any $N > 0$ that satisfies $(P^N)_{ij} > 0$ for all $i, j \in V$ also satisfies $(Q^N)_{ij} > 0$ for all $i, j \in V$. So, Q is mixing and thus, irreducible. By the Perron-Frobenius theorem [7], Q has a positive eigenvector v_Q with corresponding eigenvalue $\lambda_Q > 0$, which we call the Perron eigenvalue, such that $\lambda_Q > \lambda_i$ for all other eigenvalues λ_i.

With this definition of Q, we can rewrite (3.33) as

$$\mu_{k+1}(v) = \sum_{u\in V} \mu_k(u) Q_{u,v} = (\mu_k Q)(v).$$

This implies

$$\mu_k = \mu_{k-1} Q = \ldots = \mu_0 Q^k. \tag{3.34}$$

Lemma 3.0.29. *Let (V, P) be a mixing discrete-time Markov chain with finite state space V and transition matrix P. Let Q be defined as above, and let λ be the Perron eigenvalue of Q. Let μ_k and $S(\mu_k)$ also be defined as above.*

Then,

$$\lim_{k\to\infty} \frac{S(\mu_k)}{\lambda^k} = \mu_0 \cdot v,$$

where v is the normalized right eigenvector of Q.

Proof. By Theorem 4.5.8 in [7], the mixing property of Q implies that Q is primitive.

Now, let v^* and w^* be the right and left eigenvectors for Q respectively. We want to normalize w^* and v^* so that their inner product equals 1. Define $w = \frac{w^*}{\sum_i (w^*)_i}$ and $v = \frac{v^*}{v^* \cdot w}$. The vectors w and v are scalar multiples of w^* and v^* respectively, so they are left and right eigenvectors for Q as well. For w, we have

$$wQ = \left(\frac{w^*}{\sum_i (w^*)_i}\right) Q = \frac{1}{\sum_i (w^*)_i} w^* Q = \frac{1}{\sum_i (w^*)_i} \lambda w^* = \lambda w$$

and similarly for v.

Note that our construction of v implies $v \cdot w = 1$, as desired.

Taking w and v as the left and right eigenvectors of Q with associated eigenvalue λ, we have the following equation by Theorem 4.5.12 in [7] :

$$(Q^n)_{ij} = [(v_i w_j) + \rho_{ij}(n)]\lambda^n \tag{3.35}$$

where $\rho_{ij}(n) \to 0$ as $n \to \infty$.

Now, let us consider $S(\mu_k)$, starting with the definition.

$$\begin{aligned} S(\mu_k) &= \sum_j (\mu_k)_j \\ &= \sum_j (\mu_0 Q^k)_j \text{ by (3.34)} \\ &= \sum_j \sum_i (\mu_0)_i (Q^k)_{ij} \text{ by expanding the summand} \\ &= \sum_j \sum_i (\mu_0)_i [(v_i w_j) + \rho_{ij}(k)]\lambda^k \text{ by (3.35).} \end{aligned} \tag{3.36}$$

We can divide both sides of the above equation by λ^k to obtain

$$\frac{S(\mu_k)}{\lambda^k} = \sum_j \sum_i (\mu_0)_i (v_i w_j) + \sum_j \sum_i (\mu_0)_i \rho_{ij}(k).$$

The right-hand side can be simplified as follows.

$$\begin{aligned} \frac{S(\mu_k)}{\lambda^k} &= \sum_j \left[w_j \sum_i (\mu_0)_i v_i \right] + \sum_j \sum_i (\mu_0)_i \rho_{ij}(k) \\ &= (\mu_0 \cdot v) \sum_j w_j + \sum_j \sum_i (\mu_0)_i \rho_{ij}(k) \\ &= (\mu_0 \cdot v) + \sum_j \sum_i (\mu_0)_i \rho_{ij}(k) \text{ since } \sum_i w_i = 1 \text{ by definition.} \end{aligned}$$

Note that μ_0 and v are constant for a fixed Markov chain. Since, $\lim_{k\to\infty} \rho_{ij}(k) = 0$, we have $\lim_{k\to\infty} \sum_j \sum_i (\mu_0)_i \rho_{ij}(k) = 0$. Thus, $\lim_{k\to\infty} \frac{S(\mu_k)}{\lambda^k} = \mu_0 \cdot v$. □

Lemma 3.0.30. *Under the same conditions as Lemma 3.0.29,*

$$\lim_{k\to\infty} \frac{\mu_k}{S(\mu_k)} = w. \tag{3.37}$$

Corollary 3.0.31. *Under the conditions of Lemma 3.0.30,*

$$\lim_{k\to\infty} \mathbb{P}(X_k = (v, v) | I_k) = w_v.$$

Proof of Corollary (3.0.31). Note that $\frac{\mu_k}{S(\mu_k)}$ is a vector with elements $\left(\frac{\mu_k}{S(\mu_k)}\right)_v = \mathbb{P}(X_k = (v,v)|I_k)$ for $v \in V$. By definition,

$$\left(\frac{\mu_k}{S(\mu_k)}\right)_v = \frac{\mathbb{P}(X_k = (v,v), I_k)}{\sum\limits_{w \in V} \mathbb{P}(X_k = (w,w), I_k)}.$$

Note that the denominator reduces to $\mathbb{P}(I_k)$. Thus,

$$\left(\frac{\mu_k}{S(\mu_k)}\right)_v = \frac{\mathbb{P}(X_k = (v,v), I_k)}{\mathbb{P}(I_k)} = \mathbb{P}(X_k = (v,v)|I_k).$$

The last equality follows from the formula for conditional probability. Lemma (3.0.30) completes the proof. □

Proof of Lemma (3.0.30). Consider $\frac{\mu_k}{S(\mu_k)}$ for arbitrary k. Define $C_k = \frac{S(\mu_k)}{\lambda^k}$. By (3.34), we have

$$\frac{\mu_k}{S(\mu_k)} = \frac{\mu_0 Q^k}{C_k \lambda^k} = \frac{\mu_0}{C_k}\left(\frac{1}{\lambda^k}Q^k\right). \tag{3.38}$$

Thus, for arbitrary $j \in V$,

$$\begin{aligned}
\left(\frac{\mu_k}{S(\mu_k)}\right)_j &= \left[\frac{\mu_0}{C_k}\left(\frac{1}{\lambda^k}Q^k\right)\right]_j \text{ by (3.38)} \\
&= \sum_i \frac{(\mu_0)_i}{C_k}\left(\frac{1}{\lambda^k}Q^k\right)_{ij} \qquad (3.39)\\
&= \sum_i \frac{(\mu_0)_i}{C_k}(v_i w_j + \rho_{ij}(k)) \text{ by (3.35).}
\end{aligned}$$

Observe that

$$\sum_i \frac{(\mu_0)_i}{\mu_0 \cdot v}(v_i w_j) = w_j \frac{\sum\limits_i (\mu_0)_i v_i}{\mu_0 \cdot v} = w_j. \tag{3.40}$$

Since $\lim\limits_{k\to\infty} C_k = \mu_0 \cdot v$ and $\lim\limits_{k\to\infty} \rho_{ij}(k) = 0$, the right-hand side of (3.39) converges to the left-hand side of (3.40), and so

$$\left(\frac{\mu_k}{S(\mu_k)}\right)_j \to w_j.$$

This holds for arbitrary j, so it holds for all j. Thus, $\lim\limits_{k\to\infty} \frac{\mu_k}{S(\mu_k)} = w$ as desired. □

Lemma 3.0.32. *Let (V, P, π) be a mixing discrete-time Markov chain with finite state space V, transition matrix P, and stationary distribution π. Let X_t be the state of the product chain of (V, P, π) at time t. Let A_k be the event that $X_k = (u, u)$ for* *some $u \in V$ (i.e. X_k is on the diagonal of the product chain). Let $I_k = \cap_{i=1}^k A_k$. Then*

$$\lim_{k\to\infty} \mathbb{P}(A_k|I_{k-1}) = \sum_{i,j} w_i P_{ij}^2,$$

where w is the normalized left eigenvector of the matrix Q with entries $Q_{ij} = P_{ij}^2$.

Proof. For notational simplicity, let $X_t = i$ denote the event $X_t = (i,i)$. First, we note that

$$\mathbb{P}(A_k|I_{k-1}) = \sum_i \mathbb{P}(A_k, X_{k-1} = i|I_{k-1}).$$

We can multiply the right-hand side by $\frac{\mathbb{P}(X_{k-1}=i, I_{k-1})}{\mathbb{P}(X_{k-1}=i, I_{k-1})}$ and re-order terms in the numerator and denominator to obtain

$$\mathbb{P}(A_k|I_{k-1}) = \sum_i \mathbb{P}(X_{k-1} = i|I_{k-1})\mathbb{P}(A_k|X_{k-1} = i, I_{k-1}).$$

By the Markov property, the second factor becomes $\mathbb{P}(A_k|X_{k-1} = i)$. By stationarity, $\mathbb{P}(A_k|X_{k-1} = i) = \mathbb{P}(A_2|X_1 = i)$. We can express this probability in terms of transition probabilities as

$$\mathbb{P}(A_k|X_{k-1} = i) = \sum_j P_{ij}^2.$$

Now, we have

$$\mathbb{P}(A_k|I_{k-1}) = \sum_i \left[\mathbb{P}(X_{k-1} = i|I_{k-1}) \sum_j P_{ij}^2\right] = \sum_i \sum_j \left[\mathbb{P}(X_{k-1} = i|I_{k-1}) P_{ij}^2\right]. \tag{3.41}$$

By Corollary (3.0.31), we obtain

$$\mathbb{P}(A_k|I_{k-1}) \to \sum_i \sum_j w_i P_{ij}^2.$$

□

Now, we can prove the following formula for L.

Lemma 3.0.33. *Under the conditions of Lemma 3.0.32,*

$$L = \lim_{n\to\infty} -\frac{1}{n} \log \Delta_n.$$

Proof. Note that Δ_n can be written as $\Delta_n = \mathbb{P}(A_0) \cdot \prod_{k=1}^{n-1} \mathbb{P}(A_k|I_{k-1})$, where $\mathbb{P}(A_0) = \sum_{i \in V} \pi_i^2$. Thus,

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n}\log \Delta_n &= \lim_{n\to\infty} -\frac{1}{n}\log(\mathbb{P}(A_0) \cdot \prod_{k=1}^{n-1} \mathbb{P}(A_k|I_{k-1})) \\
&= \lim_{n\to\infty} -\frac{1}{n}\log(\mathbb{P}(A_0)) + \lim_{n\to\infty} -\frac{1}{n}\log(\prod_{k=1}^{n-1} \mathbb{P}(A_k|I_{k-1})) \\
&= \lim_{n\to\infty} -\frac{1}{n}\log\left(\prod_{k=1}^{n} \mathbb{P}(A_k|I_{k-1})\right).
\end{aligned}$$

By Lemmas 3.0.28 and 3.0.32,

$$\lim_{n\to\infty} -\frac{1}{n}\log\left(\prod_{k=1}^{n} \mathbb{P}(A_k|I_{k-1})\right) = -\log \sum_{i,j} w_i P_{ij}^2.$$

Observe that

$$\begin{aligned}
-\log \sum_{i,j} w_i P_{ij}^2 &= -\log \sum_{i,j} w_i Q_{ij} \text{ by definition of } Q \\
&= -\log \sum_j (wQ)_j \\
&= -\log \sum_j (\lambda w)_j \text{ by definition of } w \\
&= -\log(\lambda \sum_j w_j) \\
&= -\log \lambda \text{ since } w \text{ is a probability vector by definition} \\
&= L \text{ by definition.}
\end{aligned}$$

□

3.0.4 Proposition 2: $L = \lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}C_n$

Lemma 3.0.34. *Let (V, P) describe a mixing discrete-time Markov chain with state space V and transition matrix P with entropy h and stationary distribution π. Let (V_n, P_n) be the n-block process on (V, P) with stationary distribution π_n. Define $\Delta_n = \sum_{u \in V_n} ((\pi_n)_u)^2$. Let $\mathbb{E}C_n$ be the expected coalescence time for the coalescing random walk on (V_n, P_n). Then, $\lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}C_n$ exists, and*

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}C_n = L(P, V).$$

Proof. From Lemma 3.0.25, we have that

$$k_1 \frac{1}{\Delta_n} \leq \mathbb{E}C_n \leq k_2 n^2 \frac{1}{\Delta_n}.$$

Taking the logarithm yields

$$\log k_1 + \log\left(\frac{1}{\Delta_n}\right) \leq \log \mathbb{E}C_n \leq \log k_2 + \log n^2 + \log\left(\frac{1}{\Delta_n}\right).$$

Dividing by n gives

$$\frac{\log k_1}{n} - \frac{\log \Delta_n}{n} \leq \frac{\log \mathbb{E}C_n}{n} \leq \frac{\log k_2}{n} + \frac{2\log n}{n} - \frac{\log \Delta_n}{n}. \tag{3.42}$$

We previously showed that $\lim_{n\to\infty} -\frac{1}{n}\log \Delta_n$ exists in Lemma 3.0.27. Observe that taking the limit as $n \to \infty$ of the leftmost and rightmost sides of (3.42) yields $\lim_{n\to\infty} -\frac{1}{n}\log \Delta_n$ in both cases. By the Squeeze Theorem, it follows that $\lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}C_n$ exists and equals $\lim_{n\to\infty} -\frac{1}{n}\log \Delta_n$.

Thus, by Lemma (3.0.33),

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}C_n = \lim_{n\to\infty} -\frac{1}{n}\log \Delta_n = L(P, V)$$

as desired. □

3.0.5 Proposition 3: Formulas for L in terms of meeting times

In this subsection, we will prove formulas for L in terms of m_n^*, the max meeting time, and $\bar{m}_n$, the mean meeting time.

Lemma 3.0.35. *Assume the conditions of Lemma 3.0.34. Assume that $|V| > 2$ and let $m_n^* = \max_{ij\in V_n} EM_{ij}$. Then, $\lim_{n\to\infty} \frac{1}{n}\log m_n^*$ exists and*

$$\lim_{n\to\infty} \frac{1}{n}\log m_n^* = L(P, V).$$

Proof. By Lemma 3.0.22 and Proposition 3.0.16, there exists a constant $k_1 > 0$ such that

$$k_1 \frac{1}{\Delta_n} \leq m_n^* \leq \mathbb{E}C_n.$$

We can take the logarithm and divide by n to obtain

$$\frac{1}{n}\log k_1 - \frac{1}{n}\log \Delta_n \leq \frac{1}{n}\log m_n^* \leq \frac{1}{n}\log \mathbb{E}C_n.$$

In Lemmas 3.0.27 and 3.0.33 we proved that the limit of the leftmost side as $n \to \infty$ exists and equals L. In Lemma 3.0.34 we proved that the limit of the rightmost side as $n \to \infty$ exists and equals L as well. Thus, by the Squeeze Theorem, $\lim_{n\to\infty} \frac{1}{n}\log m_n^*$ exists and equals L. □

Lemma 3.0.36. *Assume the conditions of Lemma 3.0.35. Define* $\bar{m}_n = \mathbb{E}_{\pi_n \times \pi_n}(\mathbb{E}M_{ij})$. *Then,* $\lim_{n\to\infty} \frac{1}{n} \log \bar{m}_n$ *exists and*

$$\lim_{n\to\infty} \frac{1}{n} \log \bar{m}_n = L(P, V).$$

Proof. Recall that $\bar{m}_n$ is the average expected meeting time taken over all pairs of $i, j \in V_n$. By Lemma 3.0.24, $\exists k_1, k_2 > 0$ such that

$$k_1 \frac{1}{\Delta_n} \leq \mathbb{E}M_{ij} \leq k_2 n \frac{1}{\Delta_n}$$

for all $i, j \in V_n$, which implies

$$k_1 \frac{1}{\Delta_n} \leq \bar{m}_n \leq k_2 n \frac{1}{\Delta_n}.$$

Taking the log and dividing by n yields

$$\frac{1}{n} \log k_1 - \frac{1}{n} \log \Delta_n \leq \frac{1}{n} \log \bar{m}_n \leq \frac{1}{n} \log k_2 + \frac{1}{n} \log n - \frac{1}{n} \log \Delta_n. \tag{3.43}$$

Lemma 3.0.33 allows us to take the limit of the leftmost and rightmost sides of (3.43) as $n \to \infty$. By the Squeeze Theorem, $\lim_{n\to\infty} \frac{1}{n} \log \bar{m}_n$ exists and equals $\lim_{n\to\infty} -\frac{1}{n} \log \Delta_n$. By Lemma 3.0.33, $\lim_{n\to\infty} \frac{1}{n} \log \bar{m}_n = L$. □

3.0.6 Theorem 2: Bounds for $L = -\log \lambda$

Let $L(V, P) = \lim_{n\to\infty} -\frac{1}{n} \log \Delta_n$, which exists by Lemma (3.0.27). In this section, we will prove

$$0 < L(V, P) \leq h$$

where h is the entropy of the Markov chain. We will also show that L = h if and only if P is a measure of maximal entropy.

First, we will prove the upper bound.

Lemma 3.0.37. *Let (V, P) describe a mixing Markov chain with finite state V and transition matrix P with entropy h. Let (V_n, P_n) be the n-block chain that arises from (V, P). Then,*

$$L(V, P) \leq h.$$

Proof. First define a function $g_n : V_n \to \mathbb{R}$ with $g_n(j) = (\pi_n)_j$ for all $j \in V_n$.

Define $\Delta_n = \sum_{j \in V_n} (\pi_n)_j^2$. We observe that $\Delta_n = \mathbb{E}_{\pi_n}(g_n)$. Thus,

$$\log \Delta_n = \log \mathbb{E}_{\pi_n}(g_n).$$

By the Shannon-McMillan-Breiman Theorem,

$$h = \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}_{\pi_n}(\log g_n).$$

The log function is concave, so by Jensen's Inequality,

$$\mathbb{E}_{\pi_n}(\log g_n) \leq \log \mathbb{E}_{\pi_n}(g_n)$$

which implies

$$L = \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}_{\pi_n}(g_n) \leq \lim_{n\to\infty} -\frac{1}{n}\mathbb{E}_{\pi_n}(\log g_n) = h.$$

□

Next we will show the lower bound, for which we will need the following lemma.

Lemma 3.0.38. *Let (V, P) be a non-trivial mixing discrete-time Markov chain with finite state space V and transition matrix P. Let (V_n, P_n) be the n-block process induced by (V, P). Let π_n be the stationary distribution of (V_n, P_n). Then there exists positive constants δ and K such that for all $u \in V_n$,*

$$(\pi_n)_u \leq Ke^{-\delta n}.$$

Proof. Since (V, P) is mixing, $\exists m > 0$ such that $(P^m)_{ij} > 0$ for all $i, j \in V$. Let $m = \inf\{t|(P^t)_{ij} > 0 \text{ for all } i, j \in V\}$. Then for any n, we can write $n = ms + r$ for non-negative integers s and $0 \leq r < m$. Let $\alpha = \max_{i,j\in V}(P^m)_{ij} < 1$ and define $\alpha_2 = \alpha^{\frac{1}{m}} > 0$. Choose $\delta = \ln\frac{1}{\alpha_2} > 0$. We now consider two cases. Either $s = 0$ or $s \geq 1$. Suppose $s = 0$. Then we can write $(\pi_n)_u$ as

$$(\pi_n)_u = \pi_{u[1]}\prod_{k=1}^{r} P_{u[k],u[k+1]}.$$

Note that $s = 0$ implies that $n = r$ for some $0 < r < m$. Thus, $m - n > 0$. In general, we have $(\pi_n)_u \leq 1$. Since $\frac{1}{\alpha_2^{m-n}} > 1$, we have

$$(\pi_n)_u < \frac{1}{{\alpha_2}^{m-n}}.$$

Multiplying the right-hand side by $\frac{\alpha_2^n}{\alpha_2^n}$ gives us

$$(\pi_n)_u < \frac{\alpha_2^n}{\alpha_2^m}.$$

Let $K = \frac{1}{\alpha_2^m}$. Note that our choice of δ implies $\alpha_2 = e^{-\delta}$. Thus,

$$(\pi_n)_u < Ke^{-\delta n}$$

as desired.

Now consider the case $s \geq 1$. We can write $(\pi_n)_u$ as

$$(\pi_n)_u = \pi_{u[1]} \prod_{k=0}^{s-1} \left[\prod_{l=1}^{m} P_{u[mk+l],u[mk+l+1]} \right] \prod_{k=0}^{r-1} P_{u[ms+k],u[ms+k+1]}.$$

Note that the first and last factors in the above equality are each bounded above by 1. Thus,

$$(\pi_n)_u \leq \prod_{k=0}^{s-1} \left[\prod_{l=1}^{m} P_{u[mk+l],u[mk+l+1]}. \right]$$

Now, we observe that $\prod_{l=1}^{m} P_{u[mk+l],u[mk+l+1]} \leq (P^m)_{u[mk+1],u[m(k+1)]}$. (The right-hand side is the sum of the expression on the left-hand side over all possible paths of length m starting at state $u[mk+1]$ and ending at $u[m(k+1)]$.) Thus,

$$(\pi_n)_u \leq \prod_{k=0}^{s-1} (P^m)_{u[mk+1],u[m(k+1)]} \leq \alpha^s = (\alpha_2^m)^s.$$

Multiplying the rightmost expression by $\frac{\alpha_2^r}{\alpha_2^r}$ gives

$$(\pi_n)_u \leq \frac{1}{\alpha_2^r} \alpha_2^{ms+r} \leq \frac{1}{\alpha_2^m} \alpha_2^n.$$

As before, let $K = \frac{1}{\alpha_2^m}$. Substitute $e^{-\delta} = \alpha_2$ to obtain

$$(\pi_n)_u \leq K e^{-\delta n},$$

as desired. $\square$

Lemma 3.0.39. *Under the conditions of Lemma 3.0.38,*

$$0 < L = \lim_{n \to \infty} -\frac{1}{n} \log(\Delta_n).$$

Proof. Since Lemma 3.0.38 holds for all $u \in V_n$, it must hold for $u^* \in V_n$ such that $(\pi_n)_{u^*} = \max_{v \in V_n} (\pi_n)_v$. Thus,

$$\max_{v \in V_n} (\pi_n)_v \leq K e^{-\delta n}.$$

Recall that $\Delta_n = E_{\pi_n}(g_n)$ (see proof Lemma 3.0.37), so Δ_n is an expectation of the stationary distribution of the n-block process. Therefore, $\Delta_n \leq \max_{v \in V_n} (\pi_n)_v$, and so

$$\Delta_n \leq K e^{-\delta n}. \tag{3.44}$$

Next, we take the natural log of both sides of (3.44) and multiply by $-\frac{1}{n}$:

$$-\frac{1}{n}\log\Delta_n \geq -\frac{1}{n}\log K + \delta.$$

Now we take the limit as $n \to \infty$. (In Lemma 3.0.27 we proved that this limit exists for the left-hand side of the above equation, and it clearly exists for the right-hand side.) We obtain

$$\lim_{n\to\infty} -\frac{1}{n}\log\Delta_n \geq \delta > 0,$$

as desired. $\square$

Lemma 3.0.40. *Under the conditions in 3.0.37, $L = h$ if and only if P is a measure of maximal entropy.*

Proof. First, we will show that if P is MME, then $L = h$. We use the following fact from [7]. Let P be MME for the graph with vertex set V. Then there exists positive constants K_1, K_2 such that for all n and for any $j \in V_n$,

$$K_1 e^{-hn} \leq (\pi_n)_j \leq K_2 e^{-hn}. \tag{3.45}$$

Define $g_n : V_n \to \mathbb{R}$ as $g_n(j) = (\pi_n)_j$. Then $\min_{i\in V_n}(\pi_n)_i \leq \mathbb{E}_{\pi_n}(g_n) \leq \max_{i\in V_n}(\pi_n)_i$. Therefore, by (3.45),

$$K_1 e^{-hn} \leq \mathbb{E}_{\pi_n}(g_n) \leq K_2 e^{-hn}.$$

But since $\mathbb{E}_{\pi_n}(g_n) = \Delta_n$, we have that

$$K_1 e^{-hn} \leq \Delta_n \leq K_2 e^{-hn}.$$

Taking the logarithm and multiplying by $-\frac{1}{n}$ yields

$$-\frac{\log(K_1)}{n} + h \geq -\frac{\log(\Delta_n)}{n} \geq -\frac{\log(K_2)}{n} + h.$$

Lastly, we take the limit as $n \to \infty$:

$$h \geq \lim_{n\to\infty} -\frac{\log(\Delta_n)}{n} \geq h.$$

Thus, $L = \lim\limits_{n\to\infty} -\frac{\log(\Delta_n)}{n} = h$ as desired.

Now, we will show that if $L = h$, then P is a measure of maximal entropy.

Recall that we defined the matrix Q such that $Q_{ij} = P_{ij}^2$. Let A be the adjacency matrix for P. Define $X = \{x = (x_j)_{j=0}^{\infty} : x_j \in V \text{ and } A_{x_j,x_{j+1}} = 1 \text{ for all } j\}$ and let $\mathcal{M}(X)$ be the set of measures on X. Define a function $f : X \to \mathbb{R}$ such that $f(x) := \log Q_{x_0x_1} = 2\log P_{x_0x_1}$. Note that f only depends on the first two states x_0 and x_1. We will now state some facts from thermodynamic formalism that will be used in our proof (see [2, pages 11-12]) for summary of facts with references to original citations).

Fact 3.0.41. *For continuous $f : X \to \mathbb{R}$, we can define a notion called the pressure of f, denoted $\mathcal{P}(f)$, which satisfies the Variational Principle:*

$$\mathcal{P}(f) = \sup\{h(\mu) + \int f d\mu\}.$$

Fact 3.0.42. *Suppose f is a function that only depends on two coordinates.*

1. *Define*
$$Q^*(i,j) = \begin{cases} 0 & \text{if } A(i,j) = 0 \\ exp[f(ij)] & \text{otherwise} \end{cases}$$
Then $\mathcal{P}(f) = \log \lambda$ where λ is the Perron eigenvalue of stoch(Q^*).

2. *There exists a unique $\mu_f \in \mathcal{M}(X)$ such that*
$$h(\mu_f) + \int f d\mu_f = \mathcal{P}(f).$$

3. *μ_f is the Markov chain that corresponds with stoch(Q^*).*

Next, we state a theorem which is proved in greater generality by Parry and Tuncel (see [2] for the original citations).

Fact 3.0.43. *Let $f, g : X \to \mathbb{R}$ be functions that only depend on two coordinates. Then the following two statements are equivalent.*

1. *$\mu_f = \mu_g$*

2. *There exists a function $k : X \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that $f = g + k - k \circ \sigma + c$.*

Lastly,

Fact 3.0.44. *μ is a measure of maximal entropy iff $\mu = \mu_0$.*

Now we can begin the proof. By Fact 3.0.41,

$$\mathcal{P}(f) = \log \lambda = -L = -h \tag{3.46}$$

where the last equality follows by assumption.

Let μ_P be the measure corresponding to (V, P). Then,

$$\begin{aligned} h(\mu_P) + \int f d\mu_P &= h(P) + \sum \pi_i P_{ij} \log Q_{ij} \\ &= h(P) + 2 \sum \pi_i P_{ij} \log P_{ij} \\ &= h(P) - 2h(P) = -h(P) \\ &= \mathcal{P}(f). \end{aligned} \tag{3.47}$$

By Fact 3.0.41, (3.47) implies that $\mu_f = \mu_P$. Define the function $g := \frac{1}{2}f = \log P_{x_0,x_1}$. We see that $Q^* = P$, and since P is a stochastic matrix already we also have stoch(Q^*) = P. By Fact 3.0.42, it follows that $\mu_g = \mu_P$. Therefore, $\mu_f = \mu_g$.

Applying Fact 3.0.43 yields

$$f = g + k - k \circ \sigma + c$$

for some function k and constant c.

Using the definition $g = \frac{1}{2}f$ and rearranging a term gives us

$$g = k - k \circ \sigma + c = 0 + k - k \circ \sigma + c.$$

Applying Fact 3.0.43 again (this time in the other direction) allows us to conclude $\mu_g = \mu_0$. By Fact 3.0.44, it follows that μ_g is MME. Since $\mu_g = P$, we have that P is MME as desired. □

4

Bernoulli Markov chains

In this section, we will discuss some results for Bernoulli n-block Markov chains. Recall that a Bernoulli Markov chain has the property that the probability of transitioning to a state j depends only on j and not on the state which the chain occupied previously. Additionally, every Bernoulli chain has a stationary distribution π, and transition probabilities can be expressed as $P_{ij} = \pi_j$. Thus, every Bernoulli chain is defined by a state space V and a stationary distribution π, so we can denote a Bernoulli chain by (V, π) and the n-block process on (V, π) by (V_n, π_n), where $V_n = V^n$ and π_n is such that $\pi_n(j) = \prod_{i=1}^{n} \pi_{j[i]}$.

First, we will provide formulas for expected hitting times and expected meeting times. These formulas allow us to derive formulas for the maximum expected hitting time and the maximum expected meeting time respectively.

Proposition 4.0.45. *Let (V, π) describe a Bernoulli Markov chain with finite state space V and stationary distribution π. Let (V_n, π_n) be the n-block process that arises from (V, π). Let p_{ij}^t be the probability of transition from state i to state j in t time steps for any $i, j \in V_n$. For $i, j \in V_n$, define $I(i, j, d) = I_{j[1,n-d]=i[d+1,n]}$. Then for every $i, j \in V_n$, the expected hitting time has the formula*

$$\mathbb{E}_i T_j = \frac{1}{(\pi_n)_j} \sum_{d=0}^{n-1} (p_{jj}^d I(j, j, d) - p_{ij}^d I(i, j, d)). \tag{4.1}$$

Proof. We begin by considering the parameter Z_{ij}, defined in Aldous-Fill [1] as

$$Z_{ij} = \sum_{t=0}^{\infty} (p_{ij}^t - (\pi_n)_j).$$

Since X_0 and X_t are independent for $t \geq n$, $p_{ij}^t = (\pi_n)_j$ for $t \geq n$. This implies that

$$Z_{ij} = \sum_{t=0}^{n-1} (p_{ij}^t - (\pi_n)_j).$$

Lemma 12 from Chapter 2 of [1] provides a relationship between Z_{ij} and $\mathbb{E}_i T_j$:

$$\mathbb{E}_i T_j = \frac{Z_{jj} - Z_{ij}}{(\pi_n)_j}.$$

So we have

$$\mathbb{E}_i T_j = \frac{1}{(\pi_n)_j} \sum_{d=0}^{n-1} (p_{jj}^d I(j,j,d) - p_{ij}^d I(i,j,d)),$$

as desired. □

We also obtained a formula for the maximum expected hitting time.

Proposition 4.0.46. *Let $v_1 = \min_{i \in V} \pi_i$. Then, under the assumptions of Proposition 4.0.45,*

$$\max_{i,j \in V_n} \mathbb{E}_i T_j = \frac{1 - v_1^n}{(v_1)^n (1 - v_1)}. \tag{4.2}$$

Proof. Let $\pi = (\pi_1, \pi_2, ...\pi_m)$ with $\pi_1 \leq \pi_2 \leq ... \leq \pi_m$, let us define j_k to be the k^{th} letter of word j ($1 \leq k \leq n$) and π_{j_k} as the probability associated with the k^{th} letter of j.

We can express the stationary probability as follows:

$$(\pi_n)_j = \prod_{k=1}^{n} \pi_{j_k}.$$

Additionally, we can now express p_{ij}^d in terms of $I(i,j,d)$ and of probabilities π_{j_k} as follows:

$$p_{ij}^d = I(i,j,d) \prod_{r=0}^{d-1} \pi_{j_{n-r}}.$$

Let us first calculate $\mathbb{E}_{i^*} T_{j^*}$ for i^* and j^* such that $i^* = l_1 l_2 l_3 ... l_n$ and $j^* = \{w\}^n$ for letters l_r and w such that $\pi_w = v_1 = min(\pi_1, \pi_2, ...\pi_m)$ and $l_r \neq w$ for $1 \leq r \leq n$.

We have

$$\begin{aligned}
\mathbb{E}_{i^*}T_{j^*} &= \frac{1}{(\pi_n)_{j^*}}\sum_{d=0}^{n-1}(p^d_{j*j*}I(j*,j*,d)-p^d_{i*j*}I(i*,j*,d))\\
&= \frac{1}{(\pi_n)_{j^*}}\sum_{d=0}^{n-1}(p^d_{j*j*})\\
&= \frac{1}{(\pi_w)^n}\sum_{d=0}^{n-1}(\pi_w)^d\\
&= \frac{1}{(\pi_w)^n}\left(\frac{1-\pi_w{}^n}{1-\pi_w}\right)\\
&= \frac{1}{(v_1)^n}\left(\frac{1-v_1{}^n}{1-v_1}\right).
\end{aligned}$$

Now let us consider $\mathbb{E}_iT_j$ for arbitrary i and j:

$$\begin{aligned}
\mathbb{E}_iT_j &= \frac{1}{\pi_j}\sum_{d=0}^{n-1}(p^d_{jj}I(j,j,d)-p^d_{ij}I(i,j,d))\\
&\le \frac{1}{(\pi_n)_j}\sum_{d=0}^{n-1}(p^d_{jj}I(j,j,d))\\
&\le \frac{1}{(\pi_n)_j}\sum_{d=0}^{n-1}(p^d_{jj})\\
&\le \prod_{k=1}^{n}\frac{1}{\pi_{j_k}}+\left(\prod_{k=1}^{n}\frac{1}{\pi_{j_k}}\right)\sum_{d=1}^{n-1}\prod_{r=0}^{d-1}\pi_{j_{n-r}}\\
&= \prod_{k=1}^{n}\frac{1}{\pi_{j_k}}+\sum_{d=1}^{n-1}\prod_{r=1}^{n-d}\frac{1}{\pi_{j_r}}\\
&\le \prod_{k=1}^{n}\frac{1}{v_1}+\sum_{d=1}^{n-1}\prod_{r=1}^{n-d}\frac{1}{v_1}\\
&= \frac{1}{(v_1)^n}+\sum_{d=1}^{n-1}\frac{1}{(v_1)^{n-d}}\\
&= \frac{1}{(v_1)^n}\sum_{d=0}^{n-1}(v_1)^d\\
&= \mathbb{E}_{i^*}T_{j^*}.
\end{aligned}$$

Since $\mathbb{E}_iT_j \le \mathbb{E}_{i^*}T_{j^*}$ for all i and j, we have

$$\max_{i,j}\mathbb{E}_iT_j = \frac{1-v_1{}^n}{(v_1)^n(1-v_1)}.$$

□

For the expected meeting time, we have the following formula.

Proposition 4.0.47. *Let (V, π) describe a Bernoulli Markov chain with finite state space V and stationary distribution π. Let (V_n, π_n) be the n-block process that arises from (V, π). Define $p = \sum_{j \in V} \pi_j^2$. Let k_{ij} be the length of the maximal common suffix of i and j. Then for every $i, j \in V_n$, the expected meeting time has the formula*

$$\mathbb{E}M_{ij} = \frac{1 - p^{n-k_{ij}}}{p^n(1-p)}. \tag{4.3}$$

If $k_{ij} \leq n-1$, then we can also express the expected meeting time with the formula

$$\mathbb{E}M_{ij} = \sum_{h=k_{ij}+1}^{n} \frac{1}{p^h}.$$

Proof. Let W_i and W_j be two independent copies of (V_n, π_n) started at states $i \in V_n$ and $j \in V_n$ respectively. Let X_t and Y_t describe the state at time $t > 0$ of W_i and W_j respectively. Given i and j, k_{ij} is fixed, so for notational simplicity, we will denote k_{ij} by k for the remainder of this proof. Let e_g be the event that $X_t = Y_t$ for $1 \leq t \leq g-1$ and $X_g \neq Y_g$, for $1 \leq g \leq n-k$. Let e_0 be the event that $X_t = Y_t$ for $1 \leq t \leq n-k$ (and therefore W_i and W_j meet on the $(n-k)^{th}$ step). The events e_g partition the outcome space, so

$$\mathbb{E}(M_{i,j}) = \sum_{g=0}^{n-k} \mathbb{P}(e_g)\mathbb{E}(M_{i,j}|e_g).$$

Since each step is independent, we have

$$\mathbb{P}(e_g) = p^{g-1}(1-p) \text{ for } 1 \leq g \leq n-k$$

and

$$\mathbb{P}(e_0) = p^{n-k}.$$

Now let us consider $\mathbb{E}(M_{i,j}|e_g)$. For $g = 0$, we have $\mathbb{E}(M_{i,j}|e_0) = n-k$ by definition. Define $T = \inf\{N : X_t = Y_t \text{ for } N \leq t \leq N+n-1\}$ and $N_0 = n+T$. (Note that N_0 has the same probability distribution as M_{ij} for $i, j \in V_n$ such that $k_{ij} = 0$.) Given e_g, we know that after the g^{th} step, the maximal common suffix of X_t and Y_t has length 0, which implies that, from the $g+1^{th}$ step onwards, the expected additional time for W_i and W_j to meet is equivalent to the expected meeting time for a pair of random walkers W_{i^*} and W_{j^*} with $k_{i^*j^*} = 0$, or $\mathbb{E}N_0$. Thus, we have $\mathbb{E}(M_{i,j}|e_g) = g + \mathbb{E}N_0$ for $1 \leq g \leq n-k$.

Substituting for $\mathbb{E}(M_{i,j}|e_g)$ and $\mathbb{P}(e_g)$ yields

$$\begin{aligned}\mathbb{E}(M_{i,j}) &= \mathbb{P}(e_0)\mathbb{E}(M_{i,j}|e_0) + \sum_{g=1}^{n-k}\mathbb{P}(e_g)\mathbb{E}(M_{i,j}|e_g)\\ &= (n-k)p^{n-k} + \sum_{g=1}^{n-k} p^{g-1}(1-p)(g+\mathbb{E}N_0).\end{aligned}$$

Using the fact that $\mathbb{E}N_0 = \mathbb{E}M_{i,j}$ for $i, j \in V$ s.t. $k_{ij} = 0$, we obtain

$$\mathbb{E}N_0 = np^n + \sum_{g=1}^{n} p^{g-1}(1-p)(g+\mathbb{E}N_0).$$

A combination of basic algebra and Lemma 4.0.48 (which follows this proof) provides

$$\mathbb{E}N_0 = \frac{1-p^n}{p^n(1-p)}. \tag{4.4}$$

After substituting (4.4) into the formula for $\mathbb{E}M_{i,j}$, we have:

$$\begin{aligned}\mathbb{E}M_{i,j} &= (n-k)p^{n-k} + \sum_{g=1}^{n-k} p^{g-1}(1-p)(g+\frac{1-p^n}{p^n(1-p)})\\ &= (n-k)p^{n-k} + \frac{1-p}{p}\sum_{g=1}^{n-k} gp^g + (1-p)\frac{1-p^n}{p^n(1-p)}\sum_{g=1}^{n-k} p^{g-1}\\ &= (n-k)p^{n-k} + \frac{1-p}{p}\left(\frac{p^{n-k+1}((n-k)p-(n-k)-1)+p}{(1-p)^2}\right) + \frac{1-p^n}{p^n}\left(\frac{1-p^{n-k}}{1-p}\right)\\ &\qquad \text{using lemma 4.0.48}\\ &= (n-k)p^{n-k} + \frac{p^n}{p^n}\left(\frac{p^{n-k}((n-k)p-(n-k)-1)+1}{1-p}\right) + \frac{1-p^{n-k}-p^n+p^{2n-k}}{p^n(1-p)}\\ &= (n-k)p^{n-k} + \frac{p^{2n-k}((n-k)p-(n-k)-1)+p^n}{p^n(1-p)} + \frac{1-p^{n-k}-p^n+p^{2n-k}}{p^n(1-p)}\\ &= (n-k)p^{n-k} + \frac{p^{n-k}(n-k)(p-1)}{1-p} + \frac{1-p^{n-k}}{p^n(1-p)}\\ &= \frac{1-p^{n-k}}{p^n(1-p)}.\end{aligned}$$

We can rewrite this formula using the following formula:

$$\frac{x^n-1}{x-1} = x^{n-1} + x^{n-2} + ... + x + 1 \text{ for } 0 < x \neq 1.$$

For $k \leq n-1$, this gives us

$$\begin{aligned}
\mathbb{E}M_{i,j} &= \frac{p^{n-k}-1}{p^n(p-1)} \\
&= \frac{1}{p^n}(p^{n-k-1} + p^{n-k-2} + ... + p + 1) \text{ for } k < n \\
&= \frac{1}{p^{k+1}} + \frac{1}{p^{k+2}} + ... + \frac{1}{p^{n-1}} + \frac{1}{p^n} \\
&= \sum_{h=k+1}^{n} \frac{1}{p^h}.
\end{aligned}$$

□

Lemma 4.0.48. *For integers $n \geq 1$ and for $p \in (0,1)$,*

$$\sum_{g=1}^{n} gp^g = \frac{p^{n+1}(np-n-1)+p}{(1-p)^2}.$$

Proof. This proof is inspired by the proof for the expected value of the geometric random variable. We begin by considering the sum $\sum_{g=1}^{n} p^g$. Using the formula for the sum of a geometric series with a common ratio $|r| < 1$, we obtain:

$$\sum_{g=1}^{n} p^g = \frac{p-p^{n+1}}{1-p}. \tag{4.5}$$

Next, we differentiate (4.5) with respect to p:

$$\begin{aligned}
\sum_{g=1}^{n} gp^{g-1} &= \frac{(1-p)(1-(n+1)p^n) - (p-p^{n+1})(-1)}{(1-p)^2} \\
&= \frac{1-p+(p-1)(n+1)p^n + p - p^{n+1}}{(1-p)^2} \\
&= \frac{p^n(np-n-1)+1}{(1-p)^2}.
\end{aligned}$$

Multiply both sides of the equation by p to obtain the desired formula. □

The following proposition easily follows from Proposition 4.0.47.

Proposition 4.0.49. *In the setting of Proposition 4.0.47, we have*

$$\max_{i,j \in V_n} \mathbb{E}M_{ij} = \frac{1-p^n}{p^n(1-p)}. \tag{4.6}$$

Proof. From Formula 4.0.47, we have

$$\mathbb{E}M_{i,j} = \frac{1 - p^{n-k_{ij}}}{p^n(1-p)}$$

Note that the denominator is independent of i and j and thus is constant for a fixed Bernoulli chain, while the numerator varies with k_{ij} and so depends on i and j. Since $n - k_{ij} > 0$ for all k_{ij} (assume $i \neq j$) and $0 < p < 1$, we know that $1 - p^{n-k_{ij}} > 0$ for all k_{ij}. The numerator $1 - p^{n-k_{ij}}$ is maximized when $p^{n-k_{ij}}$ is minimized. Since $0 < p < 1$, $p^{n-k_{ij}}$ is minimized when $n - k_{ij}$ is maximized, i.e. when $k_{ij} = 0$. Thus, $k_{ij} = 0$ maximizes $\mathbb{E}M_{i,j}$, and so we have

$$\max_{i,j\in V_n} \mathbb{E}M_{i,j} = \frac{1 - p^n}{p^n(1-p)}$$

as desired. □

It turns out that the similarity between (4.2) and (4) allows us to prove the following inequality.

Proposition 4.0.50. *Let (V, π) describe a Bernoulli Markov chain with finite state space V and stationary distribution π. Let (V_n, π_n) be the n-block process that arises from (V, π). Then,*

$$\max_{i,j\in V_n} \mathbb{E}M_{ij} \leq \max_{i,j\in V_n} \mathbb{E}_i T_j. \tag{4.7}$$

Additionally, $\max_{i,j\in V_n} \mathbb{E}M_{ij} = \max_{i,j\in V_n} \mathbb{E}_i T_j$ *if and only if π is uniformly distributed.*

Remark: An analogous inequality is proved for reversible continuous-time chains in [1].

Proof. Suppose $m = |V|$ and $\pi = (\pi_1, \pi_2, ...\pi_m)$ with $0 < \pi_1 \leq \pi_2 \leq ... \leq \pi_m < 1$. Recall that we defined $p = \sum_{i=1}^{m} \pi_i^2$. Using the fact that $\sum_{i=1}^{m} \pi_i = 1$ and $0 < \pi_1 \leq \pi_j$ for $j > 1$, we get

$$\begin{aligned} \pi_1 &= \pi_1(\pi_1 + \pi_2 + ... + \pi_m) \\ &= \pi_1^2 + \pi_1\pi_2 + ... + \pi_1\pi_m \\ &\leq \pi_1^2 + \pi_2^2 + ... + \pi_m^2 = p. \end{aligned}$$

Note that equality holds when the stationary distribution is uniform (i.e. when $\pi_i = \frac{1}{m}$ for all $1 \leq i \leq m$). The inequalities $0 < \pi_1 \leq p < 1$ imply

$$1 < \frac{1}{p} \leq \frac{1}{\pi_1},$$

and thus, for all $n \geq 1$,

$$1 < \left(\frac{1}{p}\right)^n \leq \left(\frac{1}{\pi_1}\right)^n. \tag{4.8}$$

Recall that

$$\max_{i,j} \mathbb{E}_i T_j = \frac{\pi_1^n - 1}{\pi_1^n(\pi_1 - 1)}$$
$$\max_{i,j} \mathbb{E} M_{i,j} = \frac{(p)^n - 1}{(p)^n(p-1)}.$$

We will rewrite $\max_{i,j} \mathbb{E}_i T_j$ and $\max_{i,j} \mathbb{E} M_{i,j}$ with the help of the following formula:

$$\frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + ... + x + 1 \text{ for } 0 < x \neq 1.$$

Now we have

$$\begin{aligned} \max_{i,j} \mathbb{E}_i T_j &= \frac{1}{v_1^n}(v_1^{n-1} + v_1^{n-2} + ... + v_1 + 1) \\ &= \frac{1}{v_1} + \frac{1}{v_1^2} + ... + \frac{1}{v_1^{n-1}} + \frac{1}{v_1^n}, \end{aligned}$$

and similarly,

$$\begin{aligned} \max_{i,j} \mathbb{E} M_{i,j} &= \frac{1}{p^n}(p^{n-1} + p^{n-2} + ... + p + 1) \\ &= \frac{1}{p} + \frac{1}{p^2} + ... + \frac{1}{p^{n-1}} + \frac{1}{p^n}. \end{aligned}$$

By applying (4.8), we obtain the desired inequality.

Previously, we stated that $p = v_1$ when $\mathbf{v}$ is uniformly distributed. In this case, it is clear from the symmetry of the formulas for $\max_{i,j} \mathbb{E}_i T_j$ and $\max_{i,j} \mathbb{E} M_{i,j}$ that $\max_{i,j} \mathbb{E}_i T_j = \max_{i,j} \mathbb{E} M_{i,j}$, as desired. This proves that if $\mathbf{v}$ is uniformly distributed, then $\max_{i,j} \mathbb{E} M_{i,j} = \max_{i,j} \mathbb{E}_i T_j$.

Now let us show that if $\max_{i,j} \mathbb{E} M_{i,j} = \max_{i,j} \mathbb{E}_i T_j$, then $\mathbf{v}$ is uniformly distributed.

Suppose $\max_{i,j} E M_{i,j} = \max_{i,j} E_i T_j$. Recall that

$$\max_{i,j} \mathbb{E} M_{i,j} = \sum_{k=1}^{n} \frac{1}{p^k},$$

and similarly,

$$\max_{i,j} \mathbb{E}_i T_j = \sum_{k=1}^{n} \frac{1}{{v_1}^k}$$

So,

$$\sum_{k=1}^{n} \frac{1}{p^k} = \sum_{k=1}^{n} \frac{1}{{v_1}^k}$$

$$\sum_{k=1}^{n} \left(\frac{1}{p^k} - \frac{1}{{v_1}^k} \right) = 0$$

$$\sum_{k=1}^{n} \left(\frac{{v_1}^k - p^k}{(pv_1)^k} \right) = 0$$

Consider the term $\frac{{v_1}^k - p^k}{(pv_1)^k}$. Since $p > 0$ and $v_1 > 0$, $pv_1 > 0$ and therefore $(pv_1)^k > 0$ for all k (and in particular, for $k \geq 1$). Thus, the denominator of our term is always positive. Now consider the numerator. We know that $0 < v_1 \leq p < 1$, which implies $0 < v_1^k \leq p^k < 1$ for $k \geq 1$. So, $v_1^k - p^k \leq 0$, and thus,

$$\frac{v_1^k - p^k}{(pv_1)^k} \leq 0$$

This implies that $\sum\limits_{k=1}^{n} (\frac{v_1^k - p^k}{(pv_1)^k})$ is a sum of terms each ≤ 0. The only way we can have $\sum\limits_{k=1}^{n} (\frac{v_1^k - p^k}{(pv_1)^k}) = 0$ is if $\frac{v_1^k - p^k}{(pv_1)^k} = 0$ for all $k \geq 1$. Thus,

$$v_1^k - p^k = 0.$$

Rearranging a term gives us

$$v_1^k = p^k,$$

which implies

$$|v_1| = |p|.$$

Both $v_1 > 0$ and $p > 0$, so

$$v_1 = p$$

as desired. □

Lastly, we have the following inequality that relates the entropy of the Bernoulli n-block chain with the maximum expected meeting time.

Proposition 4.0.51. *Let (V, π) describe a Bernoulli Markov chain with finite state space V and stationary distribution π. Let (V_n, π_n) be the n-block process that arises from (V, π). Let $h = h(V, \pi)$ be the entropy of (V, π). Then,*

$$\lim_{n \to \infty} \frac{1}{n} \log(\max_{i,j \in V_n} \mathbb{E}M_{i,j}) \leq h \tag{4.9}$$

Proof. Consider the n-block process on a Bernoulli Markov chain with finite state space V. Let $p = \sum_{j \in V} (\pi_j)^2$. From Proposition 4.6,

$$\max_{i,j \in V_n} \mathbb{E} M_{ij} = \frac{1-p^n}{p^n(1-p)}.$$

Using algebra, we obtain:

$$\begin{aligned}
\frac{1}{n}\log(\max_{i,j \in V_n} \mathbb{E} M_{i,j}) &= \frac{1}{n}\log(\frac{1-p^n}{p^n(1-p)}) \\
&= \frac{1}{n}[-\log(p^n) + \log(1-p^n) - \log(1-p)] \\
&= -\log(p) + \frac{1}{n}\log(1-p^n) - \frac{1}{n}\log(1-p)
\end{aligned}$$

Note that

$$\lim_{n\to\infty} \frac{1}{n}\log(1-p^n) = 0$$

and since $\log(1-p)$ is constant,

$$\lim_{n\to\infty} \frac{1}{n}\log(1-p) = 0$$

So,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log(\max_{i,j \in V_n} \mathbb{E} M_{i,j}) &= -\log(p) \\
&= -\log(\mathbb{E}_\pi(\pi_v))
\end{aligned} \tag{4.10}$$

Now consider the entropy of the Bernoulli source:

$$\begin{aligned}
h &= -\sum_{v \in A} \pi_v \log(\pi_v) \\
&= -\mathbb{E}_\pi(\log(\pi_v))
\end{aligned} \tag{4.11}$$

By Jensen's inequality, we have

$$\log(\mathbb{E}_\pi(\pi_v)) \geq \mathbb{E}_\pi(\log(\pi_v))$$

So,

$$-\mathbb{E}_\pi(\log(\pi_v)) \geq -\log(\mathbb{E}_\pi(\pi_v)). \tag{4.12}$$

Substituting (4.10) and (4.11) into (4.12) yields

$$h \geq \lim_{n\to\infty} \frac{1}{n}\log(\max_{i,j \in V_n} \mathbb{E} M_{i,j})$$

□

References

[1] David Aldous and James A. Fill, *Reversible markov chains and random walks on graphs*, URL: http://www.stat.berkeley.edu/ aldous/RWG/book.html.

[2] Mike Boyle and Karl Petersen, *Hidden Markov processes in the context of symbolic dynamics*, Entropy of hidden Markov processes and connections to dynamical systems, London Math. Soc. Lecture Note Ser., vol. 385, Cambridge Univ. Press, Cambridge, 2011, pp. 5–71. MR 2866664 (2012j:60002)

[3] Colin Cooper, Alan Frieze, and Tomasz Radzik, *Multiple random walks in random regular graphs*, SIAM J. Discrete Math. **23** (2009/10), no. 4, 1738–1761. MR 2570201 (2011a:05322)

[4] J. T. Cox, *Coalescing random walks and voter model consensus times on the torus in* $\mathbb{Z}^d$, Ann. Probab. **17** (1989), no. 4, 1333–1366. MR 1048930 (91d:60250)

[5] Rick Durrett, *Some features of the spread of epidemics and information on a random graph*, Proceedings of the National Academy of Scinces of the United States of America **107** (2010), 4491–4498.

[6] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson. MR 2466937 (2010c:60209)

[7] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995. MR 1369092 (97a:58050)

[8] Roberto Imbuzeiro Oliveira, *Mean field conditions for coalescing random walks*, URL: http://arxiv.org/abs/1109.5684, 2012.

[9] Roberto Imbuzeiro Oliveira, *On the coalescence time of reversible random walks*, Trans. Amer. Math. Soc. **364** (2012), no. 4, 2109–2128. MR 2869200

[10] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656. MR 0026286 (10,133e)